Biblioteca Universale

Daniel C. Dennett

Coscienza

Che cosa è

Daniel C. Dennett

Coscienza

Che cosa è

Traduzione di Lauro Colasanti
Illustrazioni di Paul Weiner

Coscienza
Che cosa è

A Nick, Marcel e Ray

PREFAZIONE

Nel mio primo anno d'università lessi le *Meditazioni* di Cartesio e fui subito catturato dal problema mente-corpo. Era un mistero affascinante. Come era possibile che i miei pensieri e i miei sentimenti si congiungessero all'interno dello stesso mondo con le cellule nervose e le molecole di cui era costituito il mio cervello? Ora, dopo trent'anni di riflessioni, discussioni e pubblicazioni su questo mistero, ritengo di aver compiuto qualche progresso. Penso di poter abbozzare le linee generali di una soluzione, una teoria della coscienza che fornisca le risposte (o mostri come trovare le risposte) agli interrogativi che hanno sconcertato in uguale maniera sia filosofi e scienziati sia gli uomini della strada. Il merito non è solo mio, sono stato molto aiutato. Ho avuto l'enorme fortuna di frequentare alcuni pensatori meravigliosi che mi hanno insegnato molto in modo informale, infaticabile e paziente. Li incontrerai in queste pagine, perché la storia che devo raccontare non è quella di una meditazione solitaria ma di un'odissea attraverso molti campi, e le soluzioni ai problemi emergono da un intreccio inestricabile di dialogo e disaccordo, in cui spesso impariamo di più da errori frutto d'ipotesi audaci che da caute ambiguità. Sono sicuro che ci sono ancora molti errori nella teoria che presento qui, e spero che siano audaci, perché provocheranno delle risposte migliori da parte di altri.

Le idee di questo libro sono state forgiate nel corso di molti anni, ma la loro stesura è iniziata nel gennaio 1990 ed è terminata esattamente un anno dopo, grazie alla generosità di parecchie istituzioni e all'aiuto di molti amici, studenti e colleghi. Il Zentrum für Interdisziplinäre Forschung a Bielefeld, il CREA all'École Polytechnique di Parigi e la Villa Serbelloni della Rockefeller Foundation a Bellagio mi hanno fornito le condizioni ideali per scrivere e discutere nei primi cinque mesi. La mia università, la Tufts, ha finanziato il mio lavoro tramite il

Centro per gli Studi Cognitivi e mi ha offerto la possibilità di presentarne la penultima versione in un seminario nell'autunno del 1990 che ha attratto i professori e gli studenti della Tufts University e di tutta l'area di Boston. Voglio ringraziare anche la Fondazione Kapor e la Fondazione Harkness per il finanziamento della nostra ricerca al Centro per gli Studi Cognitivi.

Parecchi anni fa Nicholas Humphrey venne a lavorare con me al Centro per gli Studi Cognitivi, e così insieme anche a Ray Jackendoff e Marcel Kinsbourne iniziammo a incontrarci regolarmente per discutere vari aspetti e problemi della coscienza. Sarebbe difficile trovare quattro approcci più differenti nei confronti della mente, ma le nostre discussioni si sono rivelate così fruttuose, e così incoraggianti, che io voglio dedicare questo libro a questi cari amici, per ringraziarli per tutto quello che mi hanno insegnato. Altri due colleghi ed amici di vecchia data hanno giocato un ruolo importante nel processo di formazione del mio pensiero; per questo sono eternamente riconoscente a Kathleen Akins e Bo Dahlbom.

Voglio ringraziare anche il gruppo dello ZIF di Bielefeld, e in particolare Peter Bieri, Jaegwon Kim, David Rosenthal, Jay Rosenberg, Eckart Scheerer, Bob van Gulick, Hans Flohr e Lex van der Heiden; il gruppo CREA di Parigi, e in particolare Daniel Andler, Pierre Jacob, Francisco Varela, Dan Sperber e Deirdre Wilson; e i «prìncipi della coscienza» che si sono uniti a Nick, Marcel, Ray e me nella Villa Serbelloni per una settimana intensamente produttiva a marzo: Edoardo Bisiach, Bill Calvin, Tony Marcel e Aaron Sloman. Grazie anche ad Edoardo e agli altri partecipanti al seminario sul *neglect* a Parma in giugno. Pim Levelt, Odmar Neumann, Marvin Minsky, Oliver Selfridge e Nils Nilsson hanno fornito preziosi consigli su vari capitoli. Voglio esprimere la mia gratitudine a Nils anche per avermi fornito la fotografia di Shakey, e a Paul Bach-y-Rita per le sue fotografie e le sue considerazioni sui congegni per la visione protesica.

Sono grato a tutti i partecipanti del seminario dello scorso autunno per l'abbondanza delle critiche costruttive, una classe che non dimenticherò mai: David Hilbert, Krista Lawlor, David Joslin, Cynthia Schossberger, Luc Faucher, Steve Weinstein, Oakes Spalding, Mini Jaikumar, Leah Steinberg, Jane Anderson, Jim Beattie, Evan Thompson, Turhan Canli, Michael Anthony, Martina Roepke, Beth Sangree, Ned Block, Jeff McConnell, Bjorn Ramberg, Phil Holcomb, Steve White, Owen Flanagan e Andrew Woodfield. Settimana dopo setti-

mana, questa banda mi ha fatto camminare sul fuoco, nel modo più costruttivo. Durante la revisione finale, Kathleen Akins, Bo Dahlbom, Doug Hofstadter e Sue Stafford hanno fornito molti suggerimenti inestimabili. Paul Weiner ha trasformato i miei schizzi grossolani in illustrazioni e disegni eccellenti.

Kathryn Wynes e più tardi Anne Van Voorhis hanno fatto uno straordinario lavoro nell'impedire a me, e al Centro, di volar via in pezzi in questi ultimi anni che sono stati molto febbrili; senza la loro efficienza e previdenza questo libro sarebbe ancora molto lontano dalla conclusione. Infine la cosa più importante: voglio esprimere il mio amore e i miei ringraziamenti a Susan, Peter, Andrea, Marvin e Brandon, la mia famiglia.

Tufts University
Gennaio 1991

1
PREAMBOLO: COME SONO POSSIBILI LE ALLUCINAZIONI?

1. *Il cervello nella vasca*

Supponiamo che degli scienziati malvagi abbiano rimosso il tuo cervello dal tuo corpo mentre dormivi e lo abbiano posto in un sistema per il mantenimento in vita in una vasca. Supponiamo poi che cerchino di farti credere che tu non sei solo un cervello in una vasca, ma sei ancora integro, vivo e vegeto, e impegnato in normali attività nel mondo reale. Questa vecchia supposizione, il cervello nella vasca, è un esperimento mentale che si trova spesso nell'armamentario di molti filosofi. È una versione aggiornata del genio maligno di Cartesio (1641),[1] un immaginario illusionista risoluto ad ingannare Cartesio assolutamente su tutto, inclusa la sua propria esistenza. Ma come Cartesio ha osservato, anche un genio maligno infinitamente potente non avrebbe potuto costringerlo a credere alla propria esistenza se non fosse esistito: *cogito ergo sum*, «penso, dunque sono». I filosofi oggi sono meno impegnati a dimostrare la propria esistenza come esseri pensanti (forse perché ritengono che Cartesio abbia risolto il problema in modo piuttosto soddisfacente) e sono più interessati a che cosa, in linea di principio, possiamo concludere dalla nostra esperienza sulla nostra natura e sulla natura del mondo in cui (apparentemente) viviamo. *Potresti* essere niente altro che un cervello in una vasca? Potresti essere *sempre* stato solo un cervello in una vasca? Se sì, riusciresti anche solo a concepire la tua spiacevole situazione (per non parlare della possibilità di confermarla)?

L'idea del cervello nella vasca è una maniera vivida per esaminare tali questioni, ma io voglio servirmi di questa vecchia supposizione per un altro uso. Voglio utilizzarla per portare alla luce alcuni fatti curiosi sulle allucinazioni, che a loro volta ci condurranno agli esordi di una teoria – una teoria empirica e scientificamente rispettabile – della coscienza umana. Nel tradi-

zionale esperimento di pensiero è ovvio che per far funzionare il trucco gli scienziati avrebbero un bel da fare nel fornire gli stimoli appropriati a tutti i nervi troncati che provenivano dai sensi, ma i filosofi hanno supposto, a beneficio dell'argomento, che per quanto il compito possa essere tecnicamente difficile è «possibile in linea di principio». Bisogna essere accorti con queste possibilità di principio. È anche possibile in linea di principio costruire una scala in acciaio inossidabile che raggiunga la Luna o scrivere, in ordine alfabetico, tutte le intelligibili conversazioni in inglese composte da meno di mille parole. Ma nessuna di queste due cose è lontanamente possibile in pratica e talvolta una *impossibilità di fatto* è teoreticamente più interessante di una *possibilità di principio*, come vedremo.

Soffermiamoci un momento a considerare, allora, quanto sarebbe scoraggiante il compito di questi scienziati malvagi. Possiamo immaginarli dapprincipio alle prese con i problemi più semplici, sui quali si basano per arrivare a quelli più complessi. Lasciamoli iniziare con un cervello in stato comatoso, tenuto in vita ma privo di qualsiasi segnale in ingresso proveniente dai nervi ottici, da quelli auditivi e somatosensori e da tutti gli altri canali afferenti (o d'ingresso) al cervello. Si suppone in genere che un cervello in questa situazione, privo di afferenze, rimarrebbe naturalmente in uno stato comatoso per sempre, e non ci sarebbe bisogno di morfina per tenerlo assopito, ma qualche indizio empirico suggerisce che potrebbe verificarsi un certo risveglio spontaneo perfino in queste terrificanti circostanze. Ritengo legittimo supporre che se ti svegliassi in un tale stato, ti troveresti in una terribile angoscia: cieco, sordo, completamente insensibile e senza alcuna percezione del tuo proprio corpo.

Non volendoti atterrire, allora, gli scienziati fanno in modo di risvegliarti iniettando della musica stereofonica (appropriatamente codificata come impulsi nervosi) nei tuoi nervi auditivi. Fanno anche in modo che i segnali normalmente provenienti dall'apparato vestibolare o orecchio interno indichino che sei sdraiato supino, ma per il resto sempre paralizzato, insensibile e cieco. Quanto supposto finora dovrebbe rientrare nei limiti della virtuosità tecnica di un prossimo futuro – forse è già possibile. Potrebbero poi andare avanti e stimolare i tratti che innervavano la tua epidermide, fornendo i segnali in ingresso che produrrebbero normalmente un piacevole tepore sulla superficie ventrale (la pancia) del tuo corpo, e (facendosi più raffinati) potrebbero stimolare i tuoi nervi epidermici dorsali (la

schiena) proprio nel modo in cui lo farebbe la punzecchiante trama di granelli di sabbia sulla schiena. «Magnifico!», ti dici. «Eccomi qui, sdraiato sulla spiaggia, paralizzato e cieco, mentre ascolto della musica piuttosto piacevole; certo, con il rischio di ustionarmi! Come sono arrivato qui e come posso fare per chiedere aiuto?»

Ma supponiamo ora che gli scienziati, avendo realizzato tutto ciò, fronteggino il problema molto più complesso di convincerti che non sei una semplice patata da spiaggia,* ma un agente capace di impegnarsi in una qualche forma di attività nel mondo. Facendo un passetto alla volta, decidono di eliminare parte della «paralisi» del tuo corpo fantasma e lasciarti infilare il dito indice destro nella sabbia. Essi cercheranno di riprodurre l'esperienza sensibile associata al movimento del dito fornendoti la retroazione cinestesica associata ai relativi segnali volitivi o motori nella parte in uscita o efferente del tuo sistema nervoso, ma devono anche far in modo di eliminare il torpore del tuo dito fantasma e fornire la stimolazione che sarebbe prodotta dal movimento della sabbia immaginaria attorno al tuo dito.

Improvvisamente, si trovano di fronte a un problema che si rivelerà presto difficilmente gestibile, perché il modo in cui sentirai la sabbia dipende dal modo in cui deciderai di muovere il dito. Il problema di calcolare la retroazione appropriata, generarla o comporla, e poi presentartela in tempo reale diventa computazionalmente intrattabile anche per il più veloce calcolatore; e se gli scienziati cattivi decidono di risolvere il problema del tempo reale precalcolando, e quindi «preconfezionando», tutte le possibili risposte per la presentazione, non faranno altro che scambiare un problema insolubile con un altro altrettanto insolubile: ci sono troppe possibilità da immagazzinare. In breve, i nostri scienziati cattivi saranno sommersi da un'*esplosione combinatoriale* non appena tenteranno di darti degli autentici poteri di esplorazione in questo mondo immaginario.[2]

I nostri scienziati sono andati a sbattere contro un muro ben conosciuto, la cui ombra si proietta sui noiosi stereotipi di ogni videogioco. Le alternative lasciate aperte all'azione devono essere strettamente – e irrealisticamente – limitate per mantenere il compito dei «rappresentatori del mondo» nei limiti della

* L'autore qui fa un gioco di parole intraducibile tra *couch potato* (che letteralmente significa «patata da divano», ma figurativamente sfaticato, indolente, che mette le radici nel divano) e *beach potato*. [*N.d.T.*]

fattibilità. Se gli scienziati non possono fare niente di meglio che convincerti che sei condannato a giocare per tutta la vita a Donkey Kong, essi sono davvero dei cattivi scienziati.

Questo problema tecnico ha una soluzione *sui generis*. È quella usata, per esempio, per facilitare il pesante compito computazionale nei simulatori di volo fortemente realistici: usare *copie* degli oggetti nel mondo simulato. È meglio usare un reale abitacolo e tirarlo e spingerlo con pistoni idraulici, invece di cercare di simulare gli innumerevoli segnali in ingresso nel fondo schiena del pilota in allenamento. In breve, c'è un solo modo per immagazzinare in maniera rapidamente accessibile un'enorme quantità di informazioni su un mondo immaginario che deve essere esplorato, ed è quello di usare un mondo *reale* (anche se in miniatura o artificiale o in gesso) per immagazzinare le sue informazioni! Far questo, ovviamente, significa «barare» se sei il genio maligno che pretende di aver ingannato Cartesio sull'esistenza di tutte le cose, ma è un modo per realizzare lo scopo con risorse non infinite.

Cartesio è stato saggio a dotare il suo immaginario genio maligno di un'*infinita* capacità di inganno. Benché il compito non sia strettamente parlando infinito, la quantità di informazione raggiungibile in breve tempo da un essere umano «in esplorazione» è spaventosamente grande. Gli ingegneri misurano il flusso d'informazione in bit per secondo o parlano della *larghezza della banda* del canale attraverso il quale scorre l'informazione. La televisione richiede una banda più ampia di quella della radio e la televisione ad alta definizione una ancora maggiore. Una olfa-tatto-televisione ad alta definizione avrebbe bisogno di una banda ancora più ampia e una olfa-tatto-televisione *interattiva* avrebbe un'ampiezza di banda astronomica, perché si ramificherebbe costantemente in migliaia di traiettorie leggermente differenti attraverso il mondo (immaginario). Getta allo scettico una moneta dubbia, e in un secondo o due solo soppesandola, graffiandola, facendola risuonare, mordendola o semplicemente osservando come il sole brilla sulla sua superficie, egli avrà consumato più bit d'informazione di quanto un supercomputer Cray possa gestire in un anno. Fare una moneta reale ma contraffatta è un gioco da ragazzi; fare una moneta simulata a partire da nient'altro che un complesso di stimolazioni nervose è fuori della portata della tecnologia umana, oggi e forse per sempre.[3]

Una conclusione che potremmo trarre da ciò è che non siamo cervelli in vasca – nel caso che ti fossi preoccupato.

Un'altra conclusione che sembrerebbe possibile trarre da ciò è che le allucinazioni forti sono semplicemente impossibili! Per allucinazione forte intendo un'allucinazione di un oggetto tridimensionale nel mondo reale che si presenta in modo concreto e persistente – a differenza di bagliori, distorsioni geometriche, aure, immagini residue, fugaci esperienze dell'arto fantasma e altre sensazioni anomale. Potremmo dire che un'allucinazione forte sarebbe costituita da un fantasma che rispondesse, che si lasciasse toccare, che resistesse con un senso di solidità, che proiettasse un'ombra, che fosse osservabile da ogni angolo in modo da potergli girare attorno e vedere l'aspetto che ha da dietro.

È possibile fare una classifica approssimativa della forza delle allucinazioni prendendo in considerazione il numero di tali caratteristiche che in esse si manifestano. I resoconti che si riferiscono ad allucinazioni *molto* forti sono rari e possiamo ora capire perché non è una coincidenza che la credibilità di tali racconti sembra, intuitivamente, essere inversamente proporzionale alla forza delle allucinazioni riferite. Noi siamo – a ragione – particolarmente scettici nei confronti dei racconti di allucinazioni molto forti perché non crediamo nei fantasmi e pensiamo che solo un fantasma reale potrebbe produrre un'allucinazione forte. (È stata proprio l'incredibile forza delle allucinazioni riferite da Carlos Castaneda in *A scuola dallo stregone* [1968] che ha fatto subito venire in mente agli scienziati che il libro fosse un'invenzione e non un'esperienza reale, malgrado egli avesse conseguito un dottorato in antropologia presso l'Università di California per la sua «ricerca» su Don Juan.)

Ma pur non sapendo se le allucinazioni *davvero* forti si verificano, non possiamo dubitare che nell'esperienza di alcuni di noi spesso rientrano delle allucinazioni convincenti e multimodali. Le allucinazioni che sono ben attestate nella letteratura di psicologia clinica consistono spesso di dettagliate fantasie ben al di là delle capacità generative dell'attuale tecnologia. Come diamine è possibile che un singolo cervello riesca a fare quello che una squadra di scienziati e animatori al calcolatore troverebbe quasi impossibile? Se tali esperienze non sono percezioni autentiche e veridiche di qualcosa reale «fuori» della mente, esse devono essere prodotte interamente all'interno della mente (o del cervello), inventate di sana pianta ma abbastanza verosimili da ingannare la stessa mente che le ha inventate.

Il modo usuale per spiegarci questo fatto è quello di supporre che le allucinazioni si verificano quando avviene una qualche bizzarra auto-stimolazione nel cervello, in particolare una stimolazione interamente generata all'interno di qualche parte o livello dei sistemi percettivi del cervello. Cartesio, nel XVII secolo, vide questa prospettiva piuttosto chiaramente nella sua discussione sull'*arto fantasma*, la stupefacente ma piuttosto normale allucinazione a cui vanno soggette le persone che subiscono un'amputazione, le quali hanno l'impressione non solo di sentire la presenza della parte amputata, ma anche pruriti, formicolii o dolori che la riguardano. (Accade spesso che le persone che hanno appena subito un'amputazione, semplicemente non credono che una gamba o un piede sia stato loro amputato fino a che non *vedono* che non c'è più, talmente vivida e realistica è la sensazione della sua presenza.) L'analogia usata da Cartesio è quella del cordone del campanello. Prima dell'avvento dei campanelli elettrici, dei citofoni o dei telefonini cellulari, i grandi palazzi erano equipaggiati con meravigliosi sistemi di fili e pulegge che permettevano di chiamare un servitore da qualunque stanza della casa. Una decisa strattonata al nastro di velluto penzolante da un buco del soffitto metteva in trazione un filo che correva su pulegge fino alla stanza della servitù dove faceva risuonare una delle tante campanelle etichettate, informando così il maggiordomo che la sua presenza era richiesta nella stanza da letto del padrone o in salotto o nella stanza del biliardo. Tali sistemi funzionavano bene, ma si prestavano a meraviglia a delle burle. Una semplice trazione sul filo del salotto in qualunque punto della sua lunghezza avrebbe spinto il maggiordomo a trotterellare ansimando verso il salotto nella sincera ma errata convinzione che qualcuno lo voleva lì – una piccola allucinazione *sui generis*. Similmente, pensò Cartesio, dal momento che le percezioni sono causate da varie e complicate catene di eventi nel sistema nervoso che conducono alla fine al centro di controllo della mente cosciente, se si potesse intervenire in qualche punto della catena (ovunque sul nervo ottico, per esempio, tra il bulbo oculare e la coscienza), l'azione diretta sui nervi produrrebbe esattamente la stessa catena di eventi prodotta da una percezione normale e veridica di qualcosa, e ciò produrrebbe, al terminale ricevente nella mente, esattamente lo stesso effetto di un'analoga percezione cosciente.

Il cervello – o una sua parte – ha inavvertitamente giocato

un tiro meccanico alla mente. Questa era la spiegazione di Cartesio dell'allucinazione da arto fantasma. Le allucinazioni da arto fantasma, benché notevolmente vivide, sono – nella nostra terminologia – relativamente deboli; consistono di dolori e pruriti disorganizzati, tutti in un'unica modalità sensoriale. Le persone amputate non vedono od odono o (per quanto ne so io) sentono l'odore del loro piede fantasma. Così qualcosa che assomiglia alla descrizione di Cartesio potrebbe essere la maniera giusta per spiegare gli arti fantasma – lasciando da parte per il momento i noti misteri sul come il cervello fisico possa interagire con la mente cosciente non fisica. Ma possiamo comprendere che perfino la parte puramente meccanica dell'interpretazione di Cartesio si rivelerebbe incapace di spiegare le allucinazioni relativamente forti; non c'è alcun modo in cui il cervello come illusionista possa immagazzinare e manipolare sufficienti informazioni false per ingannare una mente inquisitiva. Il cervello può rilassarsi e lasciare al mondo reale il compito di fornire una sovrabbondanza di informazioni vere, ma se comincia ad indurre in corto circuito i propri nervi (o a tirare i propri fili, come avrebbe detto Cartesio), otterrebbe come unico risultato solo un'allucinazione debolissima e fugace. (In modo simile, un guasto dell'asciugacapelli della vicina può provocare «neve» o «statica» o ronzii o strani lampi sul nostro apparecchio televisivo, ma se vediamo una versione modificata del telegiornale della sera, *sappiamo* che dobbiamo ricercare una causa più complessa e organizzata di un semplice asciugacapelli.)

Si potrebbe supporre che forse siamo stati troppo ingenui riguardo alle allucinazioni; forse si verificano solo le allucinazioni leggere, fugaci, tenui – quelle forti non si verificano perché non possono verificarsi! Una rapida scorsa nella letteratura sulle allucinazioni certamente suggerisce che c'è una qualche relazione inversa tra forza e frequenza – così come tra forza e credibilità. Ma la stessa scorsa ci fornisce anche un suggerimento che ci conduce ad un'altra teoria del meccanismo di produzione delle allucinazioni: una delle caratteristiche endemiche dei racconti relativi alle allucinazioni è la inusuale passività della vittima di fronte all'allucinazione. Le persone allucinate di solito non fanno altro che immobilizzarsi e strabiliarsi. Generalmente, non sentono nessun desiderio di sondare, mettere in dubbio o indagare, e non intraprendono nessun passo per interagire con le apparizioni. Sembrerebbe, per le ragioni che abbiamo appena visto, che questa passività non sia

una caratteristica inessenziale delle allucinazioni ma una precondizione necessaria al verificarsi di qualsiasi allucinazione moderatamente dettagliata e prolungata.

La passività, comunque, è soltanto un caso particolare del modo in cui le allucinazioni relativamente forti possono sopravvivere. La ragione per cui queste allucinazioni possono sopravvivere è che l'illusionista – intendendo con ciò qualunque cosa produca l'allucinazione – può «contare su» la particolare linea d'indagine intrapresa dalla vittima – nel caso della passività totale, la linea d'indagine è *nulla*. Fintanto che l'illusionista può prevedere in dettaglio la linea d'indagine che verrà effettivamente presa, deve prepararsi a mantenere in piedi l'illusione solo «nelle direzioni che la vittima osserverà». Gli scenografi cinematografici insistono nel sapere in anticipo la posizione della macchina da presa – o, se non sarà fissa, il suo esatto angolo e traiettoria – perché così sanno che devono preparare solo quel materiale che coprirà le prospettive effettivamente utilizzate. (Non per niente il *cinéma verité* fa un largo uso della camera a mano portatile che vaga liberamente.) Nella vita reale lo stesso principio è stato usato da Potemkin per economizzare sui villaggi esemplari da far ispezionare a Caterina la Grande: il suo itinerario doveva essere ferreo.

Così una soluzione al problema delle allucinazioni forti è quello di supporre che ci sia un legame tra la vittima e l'illusionista che renda possibile a quest'ultimo di costruire l'illusione *basandosi su*, e quindi essendo capace di anticipare, le intenzioni indagatrici e le decisioni della vittima. Quando l'illusionista (per esempio, un mago sul palcoscenico) è incapace di «leggere la mente della vittima» per ottenere queste informazioni, può ancora riuscire talvolta a *far imboccare* una particolare linea d'indagine tramite una sottile ma potente «pressione psicologica». Così un prestigiatore ha a disposizione vari modi conosciuti per dare alla vittima l'illusione che sta scegliendo in perfetta libertà quale carta sulla tavola esaminare, quando in realtà c'è solo una carta che può essere girata. Per ritornare al nostro precedente esperimento di pensiero, se i cattivi scienziati possono *indurre* il cervello nella vasca ad avere un insieme definito di intenzioni investigative, essi possono risolvere il problema dell'esplosione combinatoriale preparando soltanto il materiale previsto; il sistema sarà allora solo *apparentemente* interattivo. Similmente, il genio malvagio di Cartesio può mantenere in vita l'illusione con un potere non infinito se può mantenere in vita l'illusione del libero arbitrio della vit-

tima, le cui indagini sul mondo immaginario egli controlla minuziosamente.[4]

Ma c'è un modo ancora più economico (e realistico) in cui le allucinazioni potrebbero essere prodotte nel cervello, un modo che imbriglia la stessa sfrenata curiosità della vittima. Possiamo capire come funziona sfruttando un'analogia con un gioco di società.

3. *Un gioco di società chiamato psicoanalisi*

In questo gioco si chiede a una persona, che svolge il ruolo del «pollo», di uscire dalla stanza; la si avverte che durante la sua assenza un membro del gruppo sarà chiamato a raccontare un suo sogno recente. Ciò fornirà a tutti quelli che rimangono nella stanza la trama di quel sogno così che quando egli rientrerà e inizierà a fare domande alla compagnia, l'identità del sognatore rimarrà nascosta nella folla delle risposte. Il compito del pollo è quello di formulare delle domande a cui i presenti risponderanno solo con un sì o con un no fino a quando non avrà individuato in modo sufficientemente dettagliato il canovaccio del sogno; a quel punto il pollo dovrà psicanalizzare il sognatore e usando l'analisi identificarlo o identificarla.

Una volta che il pollo è uscito dalla stanza il padrone di casa spiega agli altri convenuti che non c'è nessun sogno da raccontare, che il gioco consiste nel rispondere alle domande del pollo ubbidendo alla semplice regola seguente: alle domande la cui ultima parola inizia con una lèttera compresa nella prima metà dell'alfabeto si deve rispondere in modo affermativo, a tutte le altre si deve rispondere in modo negativo, con un'eccezione: una regola superiore di non contraddizione che impedisce di dare risposte che contraddicono risposte date in precedenza. Per esempio:

D: Il sogno riguarda una *d*onna?
R: Sì.

ma se successivamente il nostro sbadato pollo domanda

D: C'è qualche personaggio femminile nel *s*ogno?
R: Sì [malgrado l'iniziale *s* dell'ultima parola, per la regola di non contraddizione].[5]

Quando il pollo ritorna nella stanza e inizia a fare domande, riceve in risposta una serie di sì e di no più o meno casuale, o in ogni caso arbitraria. I risultati sono spesso divertenti. Talvolta il tutto termina rapidamente in assurdità, come si può intuire facilmente supponendo che la domanda iniziale fosse «La trama del sogno è identica parola per parola alla trama della Divina Commedia?» oppure «C'è qualche essere animato nel sogno?». Comunque, il più delle volte si ottiene una storia bizzarra, e spesso oscena, di ridicole disavventure che si dipana per il divertimento di tutti. Quando il pollo infine decide che il sognatore – lui o lei – deve essere una persona molto malata e complessata, la compagnia gioiosamente controbatte che è il pollo stesso l'autore del «sogno». Ciò, naturalmente, non è strettamente vero. In un certo senso, il pollo è l'autore grazie alle domande che ha deciso di porre. (Nessun *altro* ha proposto di mettere i tre gorilla nella barca con la monaca.) Ma in un altro senso, il sogno è semplicemente privo di autore, e questo è il punto che mi sta a cuore. Ci troviamo di fronte a un processo di produzione di una narrazione, di accumulazione di dettagli, del tutto privo di un piano o di intenzioni da parte di un autore: un'illusione senza illusionista.

La struttura di questo gioco di società presenta una sorprendente somiglianza con una famiglia di modelli ben accreditati di sistemi percettivi. Molti ritengono che la visione umana, per esempio, non possa essere spiegata come un'elaborazione *completamente* «diretta dai dati» o «dal basso all'alto», ma ha bisogno, ai livelli più elevati, di essere integrata da una serie, «diretta dalle aspettative», di controllo delle ipotesi (o qualcosa di analogo al controllo delle ipotesi). Un altro membro della famiglia è il modello di «analisi per sintesi» della percezione che similmente suppone che le percezioni siano costruite in un processo che fa la spola in continuazione tra aspettative centralmente generate, da una parte, e conferme (o refutazioni) sorgenti dalla periferia, dall'altro (per es., Neisser, 1967). L'idea generale di queste teorie è che dopo che si è verificata una certa quantità di «pre-elaborazione» negli strati iniziali o periferici del sistema percettivo, i compiti percettivi vengono completati – gli oggetti vengono identificati, riconosciuti, catalogati – da cicli di generazione e controllo. In un ciclo del genere, i propri interessi e le proprie aspettative correnti forgiano le ipotesi che i propri sistemi percettivi devono confermare o refutare, e una rapida sequenza di generazioni di ipotesi e conferme siffatte produce il prodotto finito, il risultato, il «modello» aggiornato

del mondo del soggetto percipiente. Tali spiegazioni della percezione sono motivate da varie considerazioni, sia biologiche sia epistemologiche, e sebbene non mi azzarderei a dire che qualcuno di questi modelli sia stato dimostrato, è sicuro che gli esperimenti ispirati a tale approccio hanno dato buoni risultati. Alcuni teorici sono stati così arditi da affermare che la percezione *deve* avere questa struttura basilare.

Quale che sia il definitivo verdetto sulle teorie della percezione basate sulla generazione e controllo, noi possiamo sfruttarle per fornire una semplice e potente spiegazione delle allucinazioni. Tutto quello che abbiamo bisogno di supporre affinché un sistema percettivo normale venga posto in una modalità allucinatoria è che la parte generativa delle ipotesi nel ciclo (la parte diretta dalle aspettative) operi normalmente, mentre la parte diretta dai dati nel ciclo (la parte che conferma) entri in una serie disordinata o casuale o arbitraria di conferme e refutazioni, proprio come nel gioco di società. In altre parole, se il rumore nel canale dei dati viene arbitrariamente amplificato fino a costituire delle «conferme» e «refutazioni» (i sì e i no arbitrari del gioco di società), le aspettative, le ansie, le ossessioni e le preoccupazioni correnti della vittima spingeranno a formulare delle domande o delle ipotesi il cui contenuto rifletterà sicuramente quegli interessi, e così una «storia» si dispiegherà nel sistema percettivo senza un autore. Non dobbiamo supporre che la storia sia scritta in anticipo; non dobbiamo supporre che l'informazione sia immagazzinata o composta nella parte illusionista del cervello. Tutto quello che supponiamo è che l'illusionista entri in una modalità di conferma arbitraria e che la vittima fornisca il contenuto ponendo le domande.

Ciò fornisce nel modo più diretto possibile un legame tra lo stato emozionale della persona in preda all'allucinazione e il contenuto dell'allucinazione stessa. Il contenuto delle allucinazioni è usualmente correlato alle ansietà provate dal soggetto allucinato; questo modello delle allucinazioni tiene conto di questa caratteristica senza l'intervento di un narratore interno implausibilmente ben informato che possiede una teoria o modello della psicologia della vittima. Perché, per esempio, il cacciatore nell'ultimo giorno della stagione di caccia al daino *vede* un daino, completo di corna e coda bianca, mentre guarda una mucca nera o un altro cacciatore in un giubbotto arancione? Perché il suo interrogatore interno sta ossessivamente chiedendo: «È un daino?» e sta ottenendo ripetuti NO come

risposta, fino a che un leggero rumore nel sistema viene erroneamente amplificato in un SÌ, con catastrofiche conseguenze.

Diverse scoperte si inquadrano perfettamente in questa descrizione delle allucinazioni. Per esempio, è risaputo che le allucinazioni sono il risultato normale di una prolungata deprivazione sensoriale (vedi, per esempio, Vosberg, Fraser e Guehl, 1960). Ciò può essere spiegato plausibilmente supponendo che nella deprivazione sensoriale la parte diretta dai dati del sistema di generazione delle ipotesi e controllo, essendo priva di dati, abbassi la sua soglia per il rumore, che quindi viene amplificato in una sequenza arbitraria di segnali di conferma e refutazione, producendo, infine, allucinazioni dettagliate il cui contenuto è prodotto da null'altro che ansiose aspettative e conferme casuali. Inoltre, nella maggior parte dei resoconti, le allucinazioni sono costruite solo gradualmente (sia in condizioni di deprivazione sensoriale sia sotto l'effetto di droghe). Iniziano debolmente – per esempio, geometriche – e poi diventano più forti («oggettive» o «narrative»), e questo è esattamente ciò che ci si aspetta utilizzando questo modello (vedi, per esempio, Siegel e West, 1975).

Infine, il semplice fatto che una droga, diffondendosi nel sistema nervoso, possa produrre degli effetti così elaborati e ricchi di contenuto richiede una spiegazione – la droga *stessa* sicuramente non può «contenere la storia», sebbene qualche ingenuo sia propenso a crederlo. Non è plausibile che la droga, tramite un'attività diffusa, possa creare o anche solo mettere in moto un elaborato sistema illusionistico, mentre è facile comprendere come la droga possa agire direttamente per alzare o abbassare o sconvolgere in qualche modo arbitrario la soglia di conferma di un sistema di generazione di ipotesi.

Il modello della generazione delle allucinazioni ispirato al gioco di società potrebbe, naturalmente, spiegare anche la composizione dei sogni. Da Freud in poi non ci sono dubbi che il contenuto tematico dei sogni sia vividamente sintomatico delle pulsioni, ansietà e preoccupazioni più profonde del sognatore, ma le indicazioni che il sogno fornisce sono notoriamente ben dissimulate sotto strati di simbolismi e ambiguità. Quale tipo di processo potrebbe produrre storie che parlano così efficacemente e incessantemente agli interessi più profondi del sognatore, ma siano completamente rivestite di strati di metafore e sostituzioni? La risposta freudiana più o meno ufficiale consiste nella stravagante ipotesi di un drammaturgo interno che compone delle opere da sogno a beneficio dell'io e che

abilmente supera l'ostacolo di un censore interno camuffando il loro vero significato. (Il modello freudiano potrebbe essere chiamato il modello amletico, perché ricorda il tortuoso espediente di Amleto di far rappresentare *Trappola per topi* solo per Claudio; serve un demonio davvero astuto per sognare uno stratagemma così sottile, ma se dobbiamo dare ascolto a Freud, in tutti noi albergano dei simili virtuosi della narrazione.) Come vedremo successivamente, le teorie che postulano degli *homunculi* del genere («piccoli uomini» nel cervello) non devono sempre essere evitate, ma ogniqualvolta si chiamano in aiuto gli *homunculi* è bene che siano dei funzionari relativamente stupidi – e non come il brillante drammaturgo freudiano che ogni notte deve produrre nuove scene da sogno per ognuno di noi! Il modello che stiamo considerando elimina del tutto il drammaturgo e per fornire il contenuto conta sul «pubblico» (che svolge un ruolo analogo a quello del pollo nel gioco di società). Il pubblico non è ottuso, naturalmente, ma non deve nemmeno avere una teoria delle proprie ansietà; deve solo essere guidato da esse nel formulare le domande.

È interessante notare, tra l'altro, che una caratteristica del gioco di società che non sarebbe necessaria in un processo che produca sogni o allucinazioni è la presenza della regola superiore di non contraddizione. Poiché è presumibile che i sistemi percettivi di un individuo siano sempre impegnati nell'esplorazione di una situazione *in fieri* (piuttosto che di un *fait accompli*, la trama di un sogno già finito), le successive «conferme» contraddittorie possono essere interpretate dal meccanismo come un'indicazione di un nuovo cambiamento nel mondo, piuttosto che come una revisione nella storia conosciuta dai relatori del sogno. Il fantasma era azzurro quando lo vidi per l'ultima volta, ma improvvisamente è diventato verde; le sue mani si sono trasformate in chele, e così via. La rapidità delle metamorfosi degli oggetti nei sogni e nelle allucinazioni è una delle caratteristiche più sorprendenti di questi racconti e ciò che è ancor più sorprendente è come raramente queste avvertite metamorfosi «ci diano fastidio» mentre sogniamo. Così la fattoria nel Vermont ora si rivela essere improvvisamente una banca a Porto Rico; e il cavallo che montavo ora è un'automobile, no un motoscafo; e il mio compagno all'inizio del viaggio era mia nonna, adesso è diventato il papa. Queste cose succedono.

Questa volubilità è esattamente quello che ci aspetteremmo da un formulatore di domande attivo ma insufficientemente scettico messo di fronte ad una sequenza casuale di Sì e No.

Nello stesso tempo, anche la persistenza di alcuni temi e oggetti nei sogni, la loro riluttanza a metamorfizzarsi o scomparire, può essere spiegata limpidamente dal nostro modello. Facendo finta per il momento che il cervello usi la regola dell'alfabeto e che utilizzi l'italiano nelle sue elaborazioni, possiamo immaginare come si svolgerebbe l'interrogatorio sotterraneo che crea un sogno ossessivo:

D: Il sogno è su mio padre?
R: No.
D: È su una finestra?
R: Sì.
D: Va bene. È su mia sorella?
R: No.
D: È su mio padre?
R: No.
D: È su mio padre in finestra?
R: Sì.
D: Lo sapevo che era su mio padre! Ora, mi stava guardando?
R: Sì ...

Difficilmente si potrebbe dire che questo limitato abbozzo di teoria abbia (già) dimostrato qualcosa sulle allucinazioni o sui sogni. Però mostra – metaforicamente – come *potrebbe funzionare* una spiegazione meccanicistica di questi fenomeni e ciò costituisce un preambolo importante, giacché alcune persone sono attratte dalla tesi disfattista che la scienza non possa «in linea di principio» spiegare i vari «misteri» della mente. L'abbozzo fino a questo punto, comunque, non ha neppure sfiorato il problema della nostra *coscienza* dei sogni e delle allucinazioni. Per di più, benché abbiamo esorcizzato un *homunculus* inverosimile, l'astuto drammaturgo/illusionista che si burla della mente, abbiamo lasciato al suo posto non solo lo stupido risponditore (che presumibilmente può essere «sostituito da macchine») ma anche il formulatore di domande, il «pubblico», che appare ancora piuttosto intelligente e bisognoso di una spiegazione. Se abbiamo eliminato un birbante, non abbiamo neppure iniziato a dare una spiegazione della vittima.

Abbiamo fatto qualche progresso, comunque. Abbiamo visto come l'attenzione verso i requisiti «ingegneristici» dei fenomeni mentali possa far sorgere domande nuove, e più facilmente affrontabili, come: quali modelli della allucinazione possono evitare un'esplosione combinatoriale? Il contenuto del-

l'esperienza come potrebbe essere elaborato da processi (relativamente) stupidi e privi di comprensione? Quale specie di legami tra processi o sistemi potrebbe spiegare i risultati della loro interazione? Se dobbiamo formulare una teoria scientifica della coscienza, dobbiamo porci molte questioni del genere.

Abbiamo anche introdotto un'idea che svolgerà un ruolo centrale in ciò che segue. L'elemento chiave delle nostre spiegazioni di come siano davvero possibili i sogni e le allucinazioni è stato il principio che il cervello deve fare solo quel lavoro, qualunque esso sia, che serve a *lenire la fame epistemica*: soddisfare la «curiosità» in tutte le sue forme. Se la «vittima» è passiva o non è curiosa sull'argomento *x*, se la vittima non cerca risposte alle domande su *x*, allora non è necessario preparare materiale sull'argomento *x*. (Se non ti prude, non grattarti.) Il mondo produce un inesauribile diluvio di informazioni che bombardano i nostri sensi e quando ci concentriamo sulla quantità di ciò che ci giunge o è continuamente a nostra disposizione, spesso cadiamo nell'illusione che dobbiamo utilizzare tutto, e sempre. Ma le nostre capacità di usare le informazioni, e i nostri appetiti epistemici, sono limitate. Se i nostri cervelli possono soltanto soddisfare tutte le nostre specifiche ingordigie epistemiche non appena si presentano, noi non abbiamo motivi per lamentarci. Non saremo mai in grado di dire, infatti, che i nostri cervelli ci forniscono meno di quanto è disponibile nel mondo.

Fin qui, questo principio parsimonioso è stato solo introdotto, non convalidato. Come vedremo, il cervello non sfrutta sempre questa opzione, ma è importante non trascurare questa possibilità. Il potere che ha questo principio di dissolvere antichi enigmi non è stato generalmente riconosciuto.

4. *Anticipazioni*

Nei capitoli seguenti tenterò di spiegare che cos'è la coscienza. Più precisamente spiegherò i vari fenomeni che compongono ciò che chiamiamo coscienza, mostrando come essi siano tutti degli effetti fisici delle attività del cervello, come queste attività si siano evolute e come facciano sorgere le illusioni sui loro poteri e le loro proprietà. È molto difficile immaginare come la mente possa essere il cervello – ma non impossibile. Per riuscire a immaginarlo, bisogna avere una conoscenza davvero approfondita delle ultime scoperte scientifiche sul funzionamento del

cervello, ma, ancora più importante, bisogna imparare nuovi modi di pensare. Accumulare fatti aiuta a immaginare nuove possibilità, ma le scoperte e le teorie della neuroscienza non sono sufficienti – perfino i neuroscienziati vengono spesso sconcertati dalla coscienza. Per favorire l'espandersi della nostra immaginazione, io proporrò, insieme ai significativi fatti scientifici, una serie di storie, analogie, esperimenti di pensiero e altri espedienti ideati per fornire nuove prospettive, rompere vecchi abiti di pensiero e aiutare a organizzare i fatti in un'unica visione coerente sorprendentemente differente dal tradizionale punto di vista sulla coscienza che tendiamo ad abbracciare. L'esperimento di pensiero sul cervello nella vasca e l'analogia con il gioco della psicoanalisi erano degli esercizi di riscaldamento per il compito principale, che è abbozzare una teoria dei meccanismi biologici e *un modo di pensare* questi meccanismi che ci farà vedere come i tradizionali paradossi e misteri sulla coscienza possano essere risolti.

Nella *Parte prima*, passeremo in rassegna i problemi della coscienza e stabiliremo alcuni metodi. Ciò è più importante e difficile di quanto si potrebbe pensare. Molti problemi incontrati dalle altre teorie sono il frutto di una partenza con il piede sbagliato, di un tentativo troppo affrettato d'indovinare le risposte alle Grandi Domande. Le insolite presupposizioni di fondo della mia teoria svolgono un grande ruolo in ciò che segue, e ci permettono di *posporre* molti dei tradizionali rompicapi filosofici su cui sono inciampati altri pensatori fino a quando non avremo delineato una teoria empiricamente fondata, il che viene fatto nella *Parte seconda.*

Il modello delle Molteplici Versioni della coscienza delineato nella *Parte seconda* rappresenta un'alternativa al modello tradizionale, che io chiamo il Teatro Cartesiano. Esso richiede un ripensamento piuttosto radicale dell'idea familiare del «flusso di coscienza» e se dapprincipio sembra profondamente controintuitivo, acquista pian piano consistenza e credibilità, e mostra come affrontare fenomeni cerebrali che sono stati ignorati fino ad ora dai filosofi – e dagli scienziati. Prendendo in esame in un certo dettaglio come la coscienza può essersi evoluta, otteniamo qualche barlume su alcune caratteristiche altrimenti sconcertanti delle nostre menti. La *Parte seconda* fornisce anche un'analisi del ruolo del linguaggio nella coscienza umana e un esame delle relazioni tra il modello delle Molteplici Versioni e alcune delle più diffuse concezioni della mente e gli altri lavori teoretici nel campo multidisciplinare

della scienza cognitiva. Lungo tutto il nostro itinerario dobbiamo resistere alle allettanti semplicità del punto di vista tradizionale, fino a che non potremo ancorarci alle nuove fondamenta.

Nella *Parte terza*, armati delle nuove maniere di guidare la nostra immaginazione, possiamo confrontarci (infine) con i tradizionali misteri della coscienza: le strane proprietà del «campo fenomenico», la natura dell'introspezione, le qualità (o *qualia*) delle esperienze, la natura del sé o io e le sue relazioni con i pensieri e le sensazioni, la coscienza di creature non umane. Potremo vedere come i paradossi che mettono in difficoltà le tradizionali discussioni filosofiche di tali misteri sorgono da una *mancanza di immaginazione*, non da «intuizioni profonde», e saremo in grado di dissolvere i misteri.

Questo libro presenta una teoria che è sia empirica sia filosofica, e, poiché una teoria del genere deve rispondere a varie esigenze, ci sono due appendici che affrontano brevemente alcune sfide più tecniche che sorgono sia da una prospettiva scientifica sia da una filosofica. Nel prossimo capitolo ci rivolgeremo al problema di definire come dovrebbe essere una spiegazione della coscienza, e di chiarire se vogliamo davvero dissolvere i problemi della coscienza.

PARTE PRIMA

PROBLEMI E METODI

2

VERSO UNA SPIEGAZIONE DELLA COSCIENZA

1. *Il vaso di Pandora: la coscienza deve essere demistificata?*

> Ecco ancora degli alberi, di cui conosco le rugosità, e dell'acqua, di cui sento il sapore. E questi profumi d'erba e di stelle, la notte, in certe sere che il cuore si placa... come negherò questo mondo, di cui sento la potenza e la forza? Eppure tutta la scienza di questa terra non potrà darmi nulla che possa rendermi certo che tale mondo mi appartiene. Voi me lo descrivete e mi insegnate a classificarlo; enumerate le sue leggi, mentre, nella mia sete di sapere, ammetto che siano vere; smontate il suo congegno e la mia speranza aumenta. [...] C'era, dunque, bisogno di tanti sforzi? Le dolci linee di queste colline e la mano della sera su questo cuore agitato me ne insegnano molto di più.
>
> ALBERT CAMUS, *Il mito di Sisifo*, 1942

> Dolce è la sapienza che la Natura dona;
> Il nostro intelletto indiscreto
> Deforma l'armonioso aspetto delle cose:-
> Noi uccidiamo per dissezionare.
>
> WILLIAM WORDSWORTH, *The Tables Turned*, 1798

La coscienza umana è praticamente l'ultimo mistero che ancora sopravvive. Un mistero è un fenomeno sul quale la gente non sa – ancora – come ragionare. Ci sono stati altri grandi misteri: il mistero dell'origine dell'universo, il mistero della vita e della riproduzione, il mistero dell'evoluzione teleologica della natura, il mistero del tempo, dello spazio e della gravità. Queste non erano solo aree di ignoranza scientifica, ma di completo sconcerto e meraviglia. Noi non possediamo ancora le risposte finali a nessuna delle questioni di cosmologia e di fisica delle particelle, di genetica molecolare e della teoria evoluzionistica, ma

sappiamo come ragionare su di esse. I misteri non sono svaniti, ma sono stati domati. Non sopraffanno più i nostri sforzi di ragionare sui fenomeni, perché ora sappiamo distinguere le domande strampalate da quelle ben formulate, ed anche se scopriremo di sbagliarci completamente con qualche risposta attualmente accettata, sappiamo come dobbiamo trattare la questione per trovarne di migliori.

Con la coscienza, invece, ci troviamo ancora nella confusione più completa. La coscienza ci si presenta oggi come un tema che spesso lascia ammutoliti e perplessi anche i più sofisticati pensatori. E, come con tutti i precedenti misteri, ci sono molti che insistono – e sperano – che non ci sarà mai una demistificazione della coscienza.

I misteri sono eccitanti, dopo tutto, fanno parte delle cose che rendono la vita divertente. Nessuno apprezza il guastafeste che rivela chi è l'assassino alla gente in fila all'entrata del cinema. Una volta che il segreto è svelato, non si può più tornare nello stato di piacevole mistificazione che prima ci ammaliava. Quindi mettiamo in guardia il lettore. Se il mio tentativo di spiegare la coscienza è valido, quelli che lo leggono baratteranno il mistero con i rudimenti di una conoscenza scientifica della coscienza, uno scambio non equo per il gusto di qualcuno. Poiché alcune persone considerano la demistificazione come una dissacrazione, mi aspetto che all'inizio considerino questo libro come un atto di vandalismo intellettuale, un assalto all'ultimo santuario dell'umanità. Vorrei convincerli che non è così.

Camus suggerisce che non ha bisogno della scienza, poiché può apprendere di più dalle soffici linee delle colline e dalla mano della sera, e io non voglio mettere in discussione la sua affermazione – dati i problemi che Camus si pone. La scienza non offre una risposta a tutti i problemi. E neppure la filosofia. Ma proprio per questa ragione i fenomeni della coscienza, che sono sconcertanti di per sé indipendentemente dalle preoccupazioni di Camus, non hanno bisogno di essere protetti dalla scienza – o da questa ricerca filosofica demistificante in cui ci siamo imbarcati. Talvolta molte persone, temendo che la scienza «uccida per dissezionare», come dice Wordsworth, sono attratte dalle dottrine filosofiche che offrono una qualche garanzia contro una tale invasione. Le loro apprensioni sono giustificate, indipendentemente dalla forza o dalla debolezza di queste dottrine; in effetti *potrebbe* accadere che la demistificazione della coscienza sia una grande perdita. Io sostengo semplicemente

che ciò *non accadrà*: le perdite, se ce ne saranno, verranno ampiamente ricompensate dai vantaggi di una migliore comprensione – sia scientifica che sociale, sia teoretica che morale – offerta da una buona teoria della coscienza.

Perché, comunque, la demistificazione della coscienza *potrebbe* essere qualcosa di cui rammaricarsi? Potrebbe essere come la perdita dell'innocenza infantile, che è sicuramente una perdita, anche se è ben ricompensata. Si consideri cosa accade all'amore, per esempio, quando diventiamo più sofisticati. Possiamo capire come un cavaliere del Medioevo potesse voler sacrificare la propria vita per l'onore di una principessa alla quale non aveva mai neanche rivolto una parola – questa era un'idea particolarmente emozionante per me quando avevo circa undici o dodici anni – ma non è uno stato mentale in cui un adulto oggi possa facilmente entrare. La gente era abituata a parlare e a pensare all'amore in forme che sono oggi praticamente non disponibili – se non per i bambini e per quelli che riescono in qualche modo a cancellare le loro conoscenze di adulti. Tutti noi amiamo dire a quelli che amiamo che li amiamo e sentirci dire da loro che siamo amati – ma da adulti non siamo più così sicuri di sapere che cosa intendiamo come lo sapevamo quando eravamo bambini e l'amore era una cosa semplice.

Ci troviamo in una condizione migliore o peggiore a causa di questo cambiamento di prospettiva? Questo cambiamento, naturalmente, non è uniforme. Mentre parecchi adulti ingenui continuano a permettere ai romanzi rosa di collocarsi in cima alla lista dei libri più venduti, noi lettori sofisticati scopriamo di esserci resi abbastanza immuni dagli effetti voluti di questi libri: ci fanno ridacchiare, non piangere. O se davvero ci fanno piangere – come talvolta accade, nostro malgrado – ci imbarazza scoprire che dei trucchi così a buon mercato siano ancora in grado di influenzarci; giacché non possiamo facilmente condividere lo stato d'animo dell'eroina che si consuma chiedendosi se ha trovato il «vero amore» – come se fosse un qualche tipo particolare di sostanza (oro emotivo da distinguere dall'ottone o dal rame emotivo). Questa crescita non avviene solo nell'individuo. La nostra cultura si è fatta più sofisticata – o almeno la sofisticazione, quale che sia il suo valore, è diffusa molto più ampiamente nella cultura. Di conseguenza, il nostro concetto di amore è cambiato, e con esso la nostra sensibilità, impedendoci di avere ora alcune esperienze che emozionavano, devastavano o tonificavano i nostri antenati.

Qualche cosa di simile è avvenuto con la coscienza. Oggigiorno noi parliamo delle nostre decisioni coscienti o delle nostre abitudini inconsce, parliamo delle esperienze coscienti che noi abbiamo (a differenza, per esempio, degli sportelli di cassa automatica, che non hanno esperienze del genere) – ma non siamo più tanto sicuri di sapere cosa intendiamo quando diciamo queste cose. Mentre ci sono ancora dei pensatori pronti a sostenere che la coscienza sia un qualcosa di prezioso e autentico (come l'amore, come l'oro), una cosa che è nello stesso tempo «ovvia» e molto, molto speciale, sta diffondendosi il sospetto che ciò sia un'illusione. Forse i vari fenomeni che cospirano a creare il senso di un singolo e misterioso fenomeno non hanno un'unità ultima ed essenziale così come non la hanno i vari fenomeni che contribuiscono a creare il senso che l'amore sia una cosa semplice.

Paragoniamo l'amore e la coscienza con due altri fenomeni piuttosto differenti, le malattie e i terremoti. Anche i nostri concetti di malattia e di terremoto hanno subito una sostanziale revisione nelle ultime centinaia di anni, ma le malattie e i terremoti sono fenomeni in larga parte (sebbene non interamente) indipendenti dai nostri concetti di essi. Cambiare opinione sulle malattie non fa, di per sé, scomparire le malattie o non le fa diventare meno frequenti, sebbene provochi dei cambiamenti nella medicina e nella salute pubblica che alterano le modalità di manifestazione delle malattie. Similmente, anche i terremoti potranno un giorno diventare parzialmente dominabili, o almeno prevedibili, dagli uomini, ma non si può dire che l'esistenza dei terremoti sia influenzata dalle nostre attitudini verso di essi o dai concetti che abbiamo di essi. Con l'amore è differente. Non è più possibile per una persona sofisticata «innamorarsi» in una delle forme una volta possibili – semplicemente perché non può credere in *quelle* forme di innamoramento. Non è più possibile per me, per esempio, avere una schietta cotta adolescenziale – a meno che non «regredisca all'adolescenza» e dimentichi o abbandoni gran parte di quello che penso di sapere. Fortunatamente, io posso credere in altri tipi di amore, ma se non ci fossero stati? L'amore è uno di quei fenomeni che *dipendono dai loro concetti* – accettiamo questa semplificazione per il momento. Ce ne sono altri: il denaro ne è un chiaro esempio. Se tutti dimenticassero che cosa è il denaro, il denaro non esisterebbe più; ci sarebbero cumuli di pezzetti di carta stampata, dischetti di metallo coniati, registrazioni computerizzate di pareggi di bilancio, edifici bancari di marmo e

granito – ma niente denaro: niente inflazione o deflazione, niente tassi di scambio o interesse, nessun *valore monetario*. Svanirebbe proprio quella proprietà di quei pezzetti di carta variamente stampati che spiega – come nient'altro potrebbe – i loro tragitti di mano in mano sulla scia di varie azioni e scambi.

Secondo il punto di vista che illustrerò in questo libro, la coscienza, come l'amore e il denaro, è un fenomeno che dipende effettivamente e in una misura inaspettata dai concetti ad essa associati. Benché abbia, come l'amore, un'elaborata base biologica, deve, come il denaro, alcune delle sue più significanti caratteristiche alla cultura, dal momento che esse non sono semplicemente inerenti, in qualche modo, alla struttura fisica delle sue manifestazioni. Così se io ho ragione, e se riuscirò a demolire alcuni di quei concetti, minaccerò di estinzione tutti quei fenomeni della coscienza che da loro dipendono. Stiamo forse per entrare nel periodo postcosciente della concettualizzazione umana? Non è forse qualcosa da temere? È anche soltanto concepibile?

Se il concetto di coscienza dovesse «cadere nel dominio della scienza», che cosa accadrebbe al senso che attribuiamo ai concetti di agente morale e di libero arbitrio? Se l'esperienza cosciente fosse «ridotta» in qualche modo a semplice materia in movimento, che cosa accadrebbe al nostro apprezzamento dell'amore e del dolore e dei sogni e della gioia? Se gli esseri umani coscienti fossero «soltanto» oggetti materiali animati, come potremmo giudicare moralmente giusto o sbagliato ciò che a loro viene fatto? Queste sono alcune delle paure che alimentano la resistenza e distraggono l'attenzione di coloro che sono messi di fronte ai tentativi di spiegare la coscienza.

Io credo che tali paure siano ingiustificate, sebbene non sia ovvio accorgersene. Incrementano la posta in gioco nella sfida che sta per cominciare tra teorie e argomenti. Ci sono degli argomenti molto potenti, che prescindono da queste paure, schierati contro il tipo di teoria materialistica e scientifica che proporrò; e riconosco che spetta a me dimostrare non solo che questi argomenti siano fallaci, ma anche che una diffusa accettazione del mio punto di vista sulla coscienza non avrebbe, in ogni caso, conseguenze così terrificanti. (E se avessi scoperto invece che avrebbe conseguenze terribili? Che cosa avrei fatto in tal caso? Non avrei scritto *questo* libro, ma, oltre a ciò, non so proprio cosa avrei fatto.)

Cerchiamo, comunque, di essere ottimisti e ricordiamoci di cosa è accaduto in seguito alle precedenti demistificazioni. Non

c'è stata nessuna diminuzione del meraviglioso; al contrario, ci siamo trovati di fronte a bellezze più profonde e a visioni più splendenti della complessità dell'universo di quelle che i protettori del mistero avessero mai concepito. La «magia» delle precedenti visioni era per lo più una copertura per una banale mancanza d'immaginazione, un fastidioso inganno religiosamente inglobato nel concetto di un *deus ex machina*. Focose divinità che guidano carri dorati attraverso i cieli sembrano banali personaggi di noiosi fumetti a confronto delle affascinanti stranezze della cosmologia contemporanea; e gli intrecci ricorsivi del meccanismo del DNA fanno sembrare l'*élan vital* più o meno così interessante come la terrificante Kryptonite di Superman. Quando avremo compreso la coscienza, quando non ci sarà più mistero, la coscienza apparirà differente, ma ci sarà ancora bellezza e uno spazio ancora più grande per il timore reverenziale.

2. *Il mistero della coscienza*

In che cosa consiste, dunque, il mistero? Che cosa potrebbe essere più ovvio o certo per ciascuno di noi del fatto che egli o ella è un soggetto cosciente dell'esperienza, un soggetto che gode di percezioni e di sensazioni, che soffre per il dolore, che concepisce idee e che consciamente delibera? Ciò sembra innegabile, ma cosa mai può essere la coscienza in se stessa? Come possono dei corpi fisici viventi nel mondo fisico produrre tali fenomeni? Questo è il mistero.

Il mistero della coscienza ha diverse maniere di presentarsi e recentemente mi ha colpito di nuovo con una forza particolare una mattina mentre leggevo un libro seduto su una sedia a dondolo. Evidentemente avevo appena sollevato gli occhi dal libro e stavo fissando ciecamente la finestra, perso nei miei pensieri, quando la bellezza dell'ambiente circostante mi distrasse dalle mie cogitazioni. In quel giorno di una precoce primavera, un fascio verde-dorato di luce solare penetrava dalla finestra e attraverso un velo di germogli verdi era ancora possibile vedere le migliaia di rami e ramoscelli dell'acero del mio giardino: il tutto formava un intricato ed elegante merletto. Il vetro della finestra è antico ed è attraversato da una corrugazione appena rilevabile; mentre mi dondolavo avanti e indietro questa imperfezione del vetro provocava un'onda di saltelli sincronizzati che si spostava avanti e indietro nel fascio dei

rami, era un movimento regolare sovrapposto con notevole chiarezza allo scintillio più caotico dei rami e ramoscelli nel vento.

Poi notai che questo metronomo visivo nei rami dell'albero era legato al ritmo del concerto grosso di Vivaldi che stavo ascoltando come «musica di sottofondo» per la mia lettura. Dapprincipio pensai che, ovviamente, dovevo aver sincronizzato inconsciamente il mio dondolio con la musica – proprio come si può inconsciamente battere il tempo con il piede – ma le sedie a dondolo hanno in realtà una gamma piuttosto limitata di frequenze di dondolio facilmente mantenibili, così probabilmente la sincronia era soprattutto il frutto di una coincidenza, solo leggermente semplificata da una mia inconscia preferenza per la precisione, per il ritmo.

La mia mente saltò per un attimo a qualche processo cerebrale confusamente immaginato che potesse spiegare come noi inconsciamente aggiustiamo il nostro comportamento, compreso quello dei nostri occhi e delle nostre facoltà che dirigono l'attenzione, al fine di «sincronizzare» il «sonoro» con «le immagini», ma queste riflessioni furono a loro volta interrotte da un'improvvisa consapevolezza. *Quello che io stavo facendo* – l'interazione di esperienze e pensieri che ho appena descritto dal mio privilegiato punto di vista in prima persona – era molto più difficile da descrivere di quei processi che dietro le quinte, inconsciamente, sicuramente *avvenivano in me* ed erano in qualche modo le cause di quel che stavo facendo. Era relativamente facile immaginare qualche meccanismo sensato dietro le quinte; ciò che era incredibilmente sconcertante era quello che avveniva «nella prima poltrona centrale», «nella luce della ribalta». Il mio pensiero cosciente, e specialmente la gioia che provavo per quella combinazione di luce solare, di solari violini di Vivaldi, di rami ondulanti – più il piacere delle mie riflessioni su ciò – come poteva *tutto questo* essere solo qualcosa di fisico che avveniva nel mio cervello? Come poteva qualsiasi combinazione di avvenimenti elettro-chimici nel mio cervello equivalere in qualche modo alla piacevolezza di quelle centinaia di ramoscelli che si inginocchiavano di fronte a me a tempo di musica? Come poteva qualche evento che elaborava informazioni nel mio cervello *essere* il calore delicato della luce solare che mi accarezzava? A tale proposito, come poteva un evento nel mio cervello *essere* la mia abbozzata immagine mentale di... qualche altro evento che elabora informazioni nel mio cervello? Sembra davvero impossibile.

Sembra davvero che gli avvenimenti che *sono* i miei pensieri e le mie esperienze coscienti non possano essere avvenimenti cerebrali, ma debbano essere *qualcos'altro*, qualcosa causato o prodotto dagli avvenimenti cerebrali, senza dubbio, ma qualcosa di aggiuntivo, fatto di una sostanza differente, situato in uno spazio differente. Beh, perché no?

3. *Le attrattive della sostanza mentale*

Vediamo cosa avviene quando prendiamo questa strada innegabilmente attraente. Per prima cosa, dovresti effettuare un piccolo esperimento. Ti invito semplicemente a chiudere gli occhi, immaginare qualcosa, e poi, dopo aver creato ed esaminato attentamente un'immagine mentale, a rispondere ad alcune domande. Non leggere le domande fino a che non hai portato a termine la seguente istruzione: chiudi gli occhi e immagina, il più dettagliatamente possibile, una mucca viola.

Fatto? Ora:

1. Vedevi il fianco destro o sinistro, o il muso della tua mucca?
2. Stava ruminando o no?
3. Potevi vedere le sue mammelle?
4. Era di un viola relativamente pallido o intenso?

Se hai seguito le istruzioni, potrai probabilmente rispondere a tutte e quattro le domande senza dover inventare qualcosa retrospettivamente. Se trovi che tutte e quattro le domande siano esageratamente specifiche, probabilmente non ti sei preoccupato affatto di immaginare una mucca viola, ma hai solo distrattamente pensato: «Sto immaginando una mucca viola» o «Facciamo finta che *questo* sia immaginare una mucca viola» o hai fatto qualcosa del genere.

Adesso facciamo un secondo esercizio: chiudi gli occhi e immagina, il più dettagliatamente possibile, una mucca *gialla*.

Questa volta probabilmente puoi rispondere alle prime tre domande senza difficoltà e saprai dire qualcosa con una certa sicurezza sul tipo di giallo – oro o limone o paglierino – che copriva il fianco della mucca che hai immaginato. Ma questa volta desidero farti una domanda differente:

5. Che differenza c'è tra l'immaginare una mucca viola e l'immaginare una mucca gialla?

La risposta è ovvia: la prima mucca immaginata è viola e la seconda è gialla. Ci potrebbero essere altre differenze, ma questa è quella essenziale. Il guaio è che, essendo queste mucche solo immaginate e non reali o dipinte su qualche tela o rappresentate su qualche schermo televisivo a colori, è difficile capire che cosa potrebbe essere viola nel primo caso e giallo nel secondo. Nulla nel tuo cervello (o nel tuo globo oculare) che sia anche solo vagamente della forma di una mucca diventa viola in un caso e giallo nell'altro, e, anche se ciò avvenisse, non servirebbe a nulla, poiché l'interno del tuo cranio è nero come la pece e, inoltre, lì non hai occhi con cui vedere i colori.

Nel tuo cervello ci sono degli eventi strettamente collegati con le tue specifiche immagini mentali, così non è impossibile che in un prossimo futuro un neuroscienziato, esaminando i processi che avvengono nel tuo cervello in risposta alle mie istruzioni, possa decifrarli talmente bene da poter confermare o smentire le tue risposte alle domande da 1 a 4:

«La mucca mostrava il fianco sinistro? Sembra di sì. La configurazione di eccitamento neuronale "testa della mucca" è coerente con una presentazione nella parte alta a sinistra del quadrante visivo; inoltre, si sono potuti osservare dei segnali rivelatori di un movimento oscillatorio di un hertz che suggeriscono la ruminazione, ma non è stato possibile rilevare alcuna attività nei gruppi di rappresentazione del complesso mammario; per quanto riguarda il colore, dopo la calibratura dei potenziali evocati con i profili di rivelazione dei colori del soggetto, si può avanzare l'ipotesi che il soggetto menta sul colore: la mucca immaginata era quasi sicuramente marrone.»

Supponi che tutto ciò sia vero; supponi che la lettura scientifica della mente sia giunta a queste raffinatezze. Anche così, però, il mistero sembrerebbe inalterato: cosa è marrone quando immagini una mucca marrone? Non l'evento nel cervello che lo scienziato ha calibrato con la tua esperienza del marrone. I tipi e la posizione dei neuroni implicati, le loro connessioni con altre parti del cervello, la frequenza o l'ampiezza dell'attività, il neurotrasmettitore chimico rilasciato – nessuna di quelle proprietà è veramente la proprietà della mucca «nella tua immaginazione». E poiché tu hai effettivamente immaginato una

mucca (non stai mentendo – lo scienziato lo ha perfino confermato), una mucca immaginata è venuta all'esistenza in quel momento; qualcosa, in qualche luogo, deve aver avuto quelle proprietà in quel momento. Il mezzo in cui la mucca immaginata deve prendere forma non è costituito da sostanza cerebrale, ma da ... sostanza mentale. Cos'altro potrebbe essere?

La sostanza mentale, allora, deve essere quella «di cui son fatti i sogni» e evidentemente ha alcune proprietà notevoli. In una di queste ci siamo già imbattuti, purtroppo è una proprietà estremamente riluttante ad essere definita. In prima approssimazione, possiamo dire che la sostanza mentale *ha sempre un testimone*. Il guaio con gli eventi cerebrali, abbiamo visto, è che, per quanto essi possano «collimare» da vicino con i nostri flussi di coscienza, hanno un inconveniente palesemente fatale: *non c'è nessuno lì che li osserva*. Gli eventi che avvengono nel tuo cervello, così come quelli che avvengono nel tuo stomaco o nel tuo fegato, normalmente non sono osservati da nessun testimone, e, per di più, la presenza di un eventuale testimone non modifica il modo in cui essi avvengono. Gli eventi nella coscienza, invece, hanno «per definizione» un testimone; essi *rientrano nell'esperienza* di qualcuno *che li esperisce* ed è proprio il fatto di rientrare nell'esperienza di qualcuno che fa di loro ciò che sono: eventi *coscienti*. Un evento di cui si fa esperienza non può avvenire per conto proprio; deve costituire l'esperienza di *qualcuno*. Affinché un pensiero si presenti, qualcuno (qualche mente) deve pensarlo, affinché un dolore si verifichi qualcuno deve sentirlo, e così se una mucca viola viene ad esistere «nell'immaginazione», qualcuno deve averla immaginata.

Il guaio con i cervelli è che, quando ci guardi dentro, scopri che *non c'è nessuno in casa*. Nessuna parte del cervello è il pensatore che effettivamente pensa o lo sventurato che sente il dolore, e il cervello nel suo insieme non sembra essere un candidato migliore per questo ruolo speciale. Questo è un argomento infido. I cervelli pensano? Gli occhi vedono? O le persone vedono con i loro occhi e pensano con i loro cervelli? C'è una differenza? È solo un banale puntiglio «grammaticale» o rivela una fonte importante di confusione? L'idea di un *sé* (o di una persona o, anche, di un'anima) distinto dal cervello o dal corpo è profondamente radicata nel nostro modo di parlare, e quindi nel nostro modo di pensare.

Io ho un cervello.

Questo sembra un modo di esprimersi completamente accettabile. E non sembra significare semplicemente

Questo corpo ha un cervello (e un cuore, e due polmoni, ecc.).

o

Questo cervello ha se stesso.

È piuttosto naturale pensare a «l'io e il suo cervello» (Popper e Eccles, 1977) come due cose distinte, con differenti proprietà, per quanto strettamente possano dipendere l'uno dall'altro. Se il sé è distinto dal cervello, sembra che debba essere fatto di sostanza mentale. In latino, una cosa pensante è una *res cogitans*, il termine reso famoso da Cartesio, che ci ha dato quella che egli pensò essere un'incontrovertibile dimostrazione del fatto che, essendo palesemente una cosa pensante, egli non poteva essere il suo cervello. Eccone una versione, sicuramente dotata di una sua forza stringente:

Poi, esaminando con attenzione ciò che ero, e vedendo che potevo fingere, sì, di non avere nessun corpo, e che non esistesse il mondo o altro luogo dove io fossi, ma non perciò potevo fingere di non esserci io, perché, anzi, dal fatto stesso di dubitare delle altre cose seguiva nel modo più evidente e certo che io esistevo; laddove, se io avessi solamente cessato di pensare, ancorché tutto il resto di quel che avevo immaginato fosse stato veramente, non avrei avuto ragione alcuna per credere di esser mai esistito: ne conchiusi esser io una sostanza, di cui tutta l'essenza o natura consiste solo nel pensare, e che per esistere non ha bisogno di luogo alcuno, né dipende da cosa alcuna materiale. (*Discorso sul Metodo*, 1637)

Così abbiamo scoperto due specie di cose che si potrebbero voler fare con la sostanza mentale: la mucca viola che non è nel cervello e la cosa che effettua il pensare. Ma ci sono ancora altri poteri speciali che potremmo voler attribuire alla sostanza mentale.

Supponiamo che un'azienda vinicola decida di rimpiazzare i suoi assaggiatori di vino umani con una macchina. Un «sistema esperto», gestito da un calcolatore, per i controlli di qualità e la classificazione del vino non è poi così fuori della portata dell'attuale tecnologia. Conosciamo già a sufficienza la chimica speci-

fica che ci servirebbe per costruire i trasduttori che rimpiazzerebbero le papille gustative e i ricettori olfattivi dell'epitelio (fornendo così il «materiale grezzo» – gli stimoli in ingresso – per il gusto e l'olfatto). Ancora non sappiamo precisamente come questi segnali in ingresso si combinino e interagiscano per produrre le nostre esperienze, ma anche su questo si stanno facendo notevoli progressi. I lavori sulla visione sono molto più avanti. Le ricerche sulla visione cromatica suggeriscono che imitare tramite qualche componente meccanico la reattività specifica dell'uomo, la sua sensibilità e affidabilità, nella discriminazione dei colori sarebbe una sfida tecnologica di enorme portata, ma non un compito impossibile. Così siamo autorizzati a immaginare che utilizzando gli elaborati segnali in uscita di quei trasduttori sensoriali e il meccanismo per la loro comparazione potremmo fornire sufficiente materiale a delle routine di classificazione, descrizione e valutazione. Versiamo il campione di vino nell'imbuto e, dopo pochi minuti o poche ore, il sistema stamperà un'analisi chimica dettagliata e un commento del tipo: «Un Pinot esuberante e vellutato, benché privo di corpo». Una macchina del genere potrebbe perfino dare risultati più coerenti e affidabili di quelli degli assaggiatori umani in tutte le prove ragionevoli che i produttori di vino volessero escogitare, ma sicuramente, per quanto possa essere «sensibile» e «discriminante», sembra che un tale sistema non potrebbe mai avere, e mai gustare, ciò che *noi* abbiamo, e gustiamo, quando assaggiamo un vino.

Questo è effettivamente così ovvio? Stando alle varie teorie raggruppate sotto l'etichetta di *funzionalismo*, riuscire a riprodurre l'*intera* «struttura funzionale» del sistema cognitivo di un assaggiatore umano di vino (inclusi la sua memoria, i suoi scopi, le sue avversioni innate, ecc.) significa riprodurre anche tutte le proprietà mentali che lo riguardano, incluso quel godimento, quella delizia, quel gusto che rende una buona bevuta qualcosa che molti di noi *apprezzano*. In linea di principio, dice il funzionalista, non c'è differenza tra un sistema fatto di molecole organiche e uno fatto di silicio, fintanto che *svolgono lo stesso compito*. I cuori artificiali non devono essere fatti di tessuto organico, e neppure i cervelli artificiali – almeno in linea di principio. Se riesci a riprodurre nei chip al silicio tutte le funzioni di controllo di un cervello di un assaggiatore umano di vino, avrai riprodotto *ipso facto* anche il godimento.

Una qualche versione di funzionalismo potrà trionfare alla fine (in effetti questo libro difende un tipo di funzionalismo), ma

sicuramente sembra impossibile di primo acchito. Sembra che nessuna mera macchina, non importa quanto meravigliosamente imiti i processi cerebrali dell'assaggiatore umano di vino, possa essere in grado di apprezzare un vino, o una sonata di Beethoven o una partita di pallacanestro. Per poter apprezzare qualcosa serve la coscienza – proprio ciò che manca a qualsiasi mera macchina. Ma naturalmente il cervello *è* una sorta di macchina, un organo come il cuore o i polmoni o i reni, che in definitiva ha una spiegazione meccanica di tutti i suoi poteri. Sembra che proprio questo ci obblighi a dire che non è il cervello che apprezza; la responsabilità (o il privilegio) di *questo atto* spetta alla mente. La riproduzione del meccanismo del cervello in una macchina fatta di silicio non provocherebbe, allora, un reale apprezzamento, ma nel migliore dei casi solo un'illusione o un simulacro di apprezzamento.

Così la mente cosciente non è solo il luogo in cui si trovano i colori e gli odori presentati ad un testimone e non è solo la cosa che pensa. È anche il luogo dove avvengono gli apprezzamenti. È l'arbitro ultimo sul perché qualcosa è importante per noi. Forse, in un certo senso, ciò segue anche dal fatto che la mente cosciente è anche ritenuta essere la fonte delle nostre azioni intenzionali. È conforme a ragione – non è così? – che se *fare qualcosa che è importante per noi* dipende dalla coscienza, anche *questa importanza* (questo gioire, apprezzare, soffrire, preoccuparsi) dovrebbe dipendere dalla coscienza. Se un sonnambulo «inconsciamente» provoca un danno, non viene ritenuto responsabile perché in un senso importante *egli* non lo ha provocato; i movimenti del suo corpo sono inseparabilmente intrecciati alla catena causale che ha condotto al danno, ma essi non costituiscono una sua *azione*, così come non la costituirebbero se egli avesse provocato un danno semplicemente cadendo dal letto. La mera complicità corporale non basta a produrre un'azione intenzionale, né basta una complicità corporale *sotto il controllo delle strutture cerebrali*, giacché il corpo di un sonnambulo è chiaramente sotto il controllo delle strutture del suo cervello. Quello che deve essere aggiunto è la coscienza, l'ingrediente speciale che trasforma un mero *avvenimento* in un'*azione*.[1]

Non è colpa del Vesuvio se la sua eruzione uccide i tuoi cari, e mostrare del risentimento (Strawson, 1962) o disprezzarlo non costituiscono delle opzioni disponibili – a meno che non riesci in qualche modo a convincerti che il Vesuvio, contrariamente all'opinione attualmente accettata, sia un agente cosciente. In effetti, è strano quanto possa confortare il nostro

dolore entrare in quegli stati d'animo in cui ce la prendiamo con la «furia» dell'uragano, malediciamo il cancro che così ingiustamente ha colpito un bambino o bestemmiamo «gli dèi». Anticamente dire che qualcosa fosse «animata», in opposizione a «inanimata», significava dire che possedeva un'anima, una mente. Rappresentarci le cose che ci influenzano potentemente come animate potrebbe essere molto di più che un semplice conforto, potrebbe essere un sofisticato trucco della nostra struttura biologica, una scorciatoia per aiutare i nostri cervelli sempre a corto di tempo ad organizzarsi e a pensare alle cose di cui bisogna preoccuparsi se si vuole sopravvivere.

Forse è vero che abbiamo una tendenza innata a trattare ogni cosa cangiante, almeno inizialmente, come se avesse un'anima (Stafford, 1983; Humphrey, 1983b, 1986), ma, per quanto quest'attitudine possa essere naturale, noi sappiamo ora che attribuire un'anima (cosciente) al Vesuvio significa spingersi troppo avanti. Dove tracciare esattamente la linea è una *vexata quaestio* sulla quale ritorneremo, ma, per quanto riguarda noi stessi, sembra che la coscienza sia esattamente ciò che ci distingue dai semplici «automi». I «riflessi» meramente corporei sono «automatici» e meccanici; possono implicare i circuiti cerebrali, ma non richiedono alcun intervento della mente cosciente. Ci sembra molto naturale pensare ai nostri corpi come a delle specie di banali marionette che «noi» controlliamo «dall'interno». Faccio sollevare la mano alla marionetta muovendo il mio dito; muovo il mio dito... come, muovendo la mia anima? Questa idea presenta dei problemi ben noti, ma non ci impedisce comunque di pensare che colga qualcosa di importante: a meno che non ci sia una mente cosciente dietro l'azione, non c'è un agente *reale* in carica. Quando pensiamo alle nostre menti in questa maniera, ci sembra di scoprire l'«io interiore», l'«io reale». Questo io reale non è il mio cervello; è quello che *possiede* il mio cervello («l'io e il *suo* cervello»). Sulla scrivania di Harry Truman nell'ufficio presidenziale alla Casa Bianca c'era un famoso cartello: «Lo scaricabarile finisce qui». Sembra che nessuna parte del cervello possa essere quella *che non può più scaricare i barili su un'altra*, la fonte ultima della responsabilità morale all'inizio della catena degli ordini.

Per riassumere, abbiamo trovato quattro ragioni per credere nella sostanza mentale. La mente cosciente, a quanto pare, non può *essere semplicemente* il cervello, o nessuna parte propria di esso, giacché nulla nel cervello potrebbe

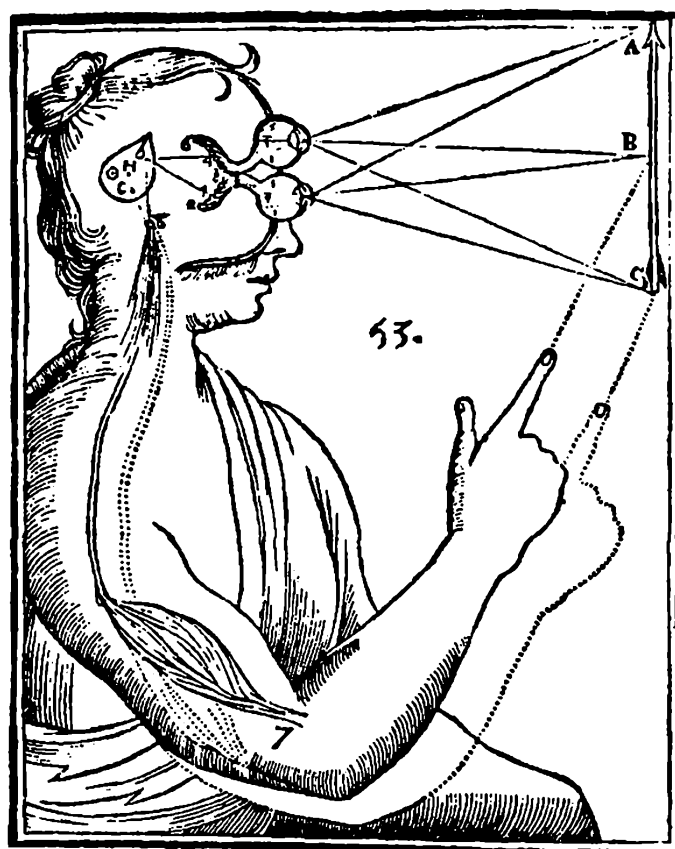
1. essere il mezzo in cui viene resa la mucca viola;
2. essere la cosa pensante, il soggetto del «*cogito, ergo sum*»;
3. apprezzare il vino, odiare il razzismo, amare qualcuno, ritenere importante qualcosa;
4. agire con responsabilità morale.

Una teoria accettabile della coscienza umana deve dare una spiegazione a questi quattro motivi stringenti che ci spingono a pensare che debba esistere una sostanza mentale.

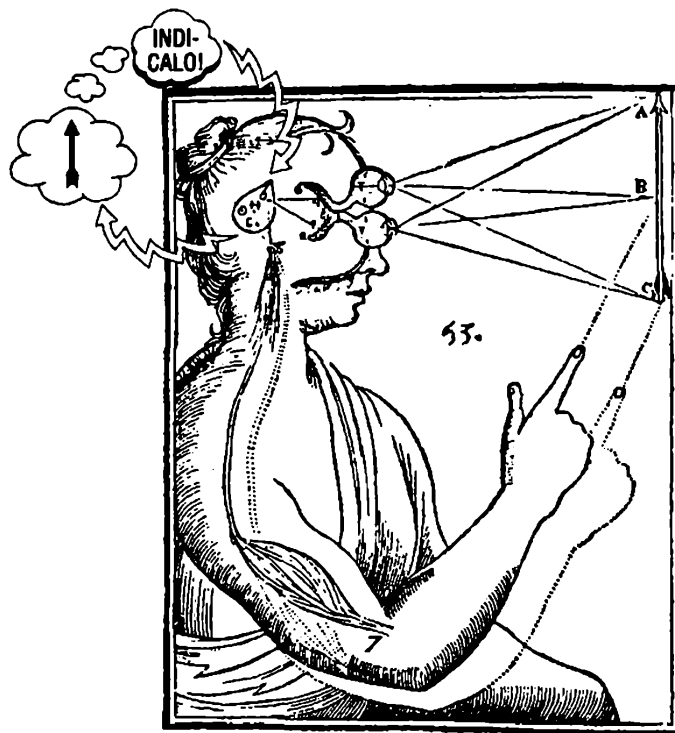
4. *Perché il dualismo è in disgrazia*

L'idea che la mente sia un'entità così separata dal cervello e composta non da materia ordinaria, ma da qualche altra sostanza speciale, viene usualmente chiamata *dualismo*. Oggigiorno esso gode, meritatamente, di una cattiva reputazione, malgrado gli argomenti convincenti appena esaminati. Fin dall'attacco ormai classico di Gilbert Ryle (1949) a ciò che egli chiamava «il dogma cartesiano dello spettro nella macchina», i dualisti sono sulla difensiva.[2] La posizione dominante, variamente espressa e sostenuta, è il *materialismo*: esiste un solo tipo di sostanza, e cioè la *materia* – la sostanza fisica di cui si occupano la fisica, la chimica e la fisiologia – e la mente è in un certo senso niente altro che un fenomeno fisico. In breve, la mente è il cervello. Secondo i materialisti, possiamo (in linea di principio!) dare una spiegazione di ogni fenomeno mentale usando gli stessi princìpi fisici, le stesse leggi e gli stessi materiali grezzi che ci bastano per spiegare la radioattività, la deriva dei continenti, la fotosintesi, la riproduzione, la nutrizione e la crescita. Uno dei compiti principali di questo libro è quello di spiegare la coscienza senza mai arrendersi al canto delle sirene del dualismo. Che c'è di errato, allora, nel dualismo? Perché è così in disgrazia?

L'obiezione usuale al dualismo era ben conosciuta dallo stesso Cartesio nel XVII secolo e sembra giusto affermare che né lui né i successivi dualisti siano mai riusciti a superarla convincentemente. Benché siano delle entità o sostanze distinte, la mente e il cervello devono tuttavia interagire; gli organi di senso corporei, attraverso il cervello, devono *informare* la mente, devono inviarle o mostrarle le percezioni o le idee o qualunque tipo di dati, e la mente dopo averci riflettuto sopra deve *dirigere* il corpo in varie azioni appropriate (incluse quelle linguisti-

che). Per tale motivo questa posizione viene spesso chiamata «interazionismo cartesiano» o «dualismo interazionista». Il sito dell'interazione nel cervello era identificato da Cartesio nella ghiandola pineale o *epifisi*: nel disegno schematico di Cartesio, è quella specie di goccia molto ingrandita nel centro della testa.

2.1

Possiamo chiarificare la difficoltà dell'interazionismo sovrapponendo al suo disegno un abbozzo del resto della sua teoria.

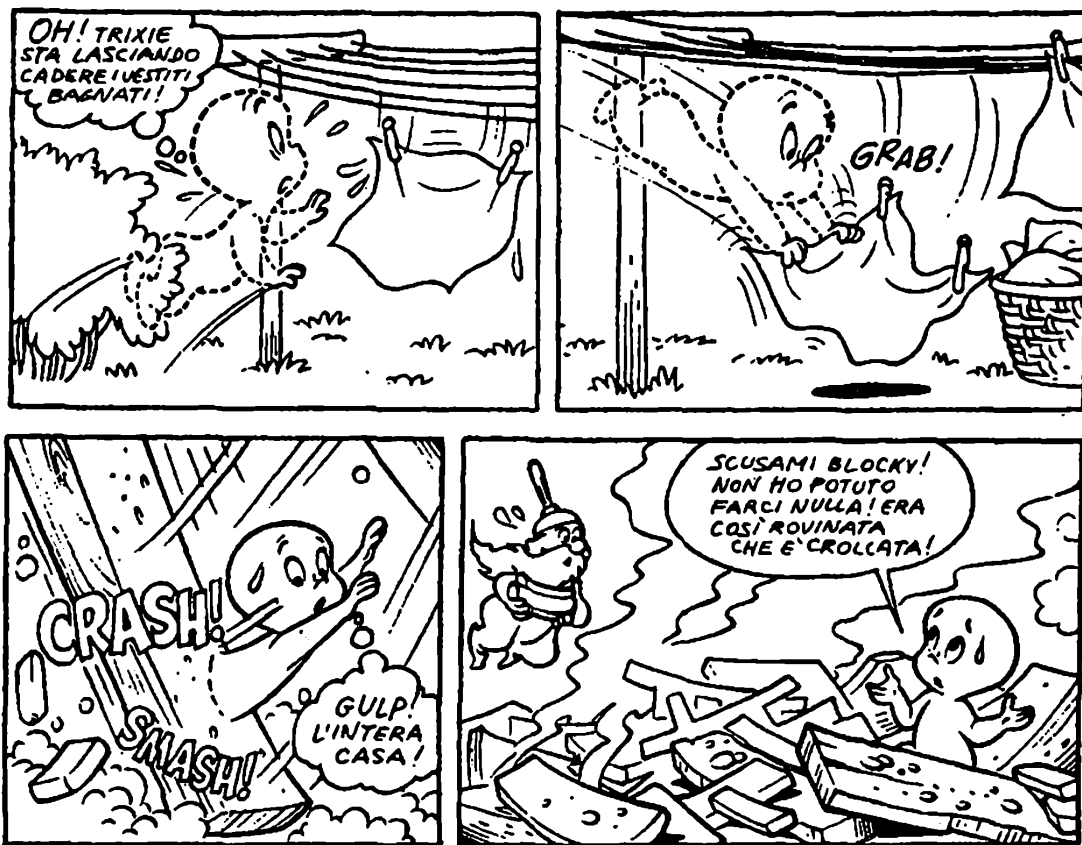
La percezione cosciente della freccia avviene solo dopo che il cervello ha trasmesso in qualche modo il suo messaggio alla mente e il dito della persona può indicare la freccia solo dopo che la mente ha inviato i suoi ordini al corpo. Ma come viene trasmessa, precisamente, l'informazione dalla ghiandola pineale alla mente? Poiché non abbiamo (per ora) la minima idea delle proprietà della sostanza mentale, non possiamo neanche immaginare (per ora) come possa essere influenzata dai processi fisici che provengono in qualche modo dal cervello, quindi ignoriamo per il momento questi segnali ascendenti e concentriamoci su quelli di ritorno, quelli che vanno dalla mente al cervello. Questi, *ex hypothesi*, non sono fisici; non sono onde elettromagnetiche o onde acustiche o raggi cosmici o fasci di particelle subatomiche. Nessuna energia o massa fisica è asso-

2.2

ciata ad essi. Come riescono, allora, ad influenzare il funzionamento delle cellule cerebrali ad essi collegate?

Un principio fondamentale della fisica prescrive che qualunque cambiamento nella traiettoria di qualunque entità fisica costituisce un'accelerazione che richiede un dispendio d'energia; in questo caso, da dove viene l'energia? Questo principio della conservazione dell'energia che spiega il perché dell'impossibilità delle macchine per «il moto perpetuo» è lo stesso principio che viene manifestamente violato dal dualismo. Questa difficoltà di accordare la fisica standard con il dualismo è stata discussa infinite volte fin dai giorni di Cartesio e viene generalmente considerata come una fatale e ineluttabile pecca del dualismo.

Come era prevedibile, sono state tentate e sviluppate varie ed ingegnose eccezioni tecniche basate su letture sofisticate della fisica implicata, ma senza provocare molte conversioni. Ma non è necessario appellarsi ad assodate leggi della fisica per mettere in risalto le difficoltà del dualismo che stiamo considerando. Si tratta della stessa incoerenza che i bambini rilevano – ma tranquillamente tollerano nella fantasia – in cose come Casper il fantasma amichevole.

Come è possibile che Casper possa *sia* volare attraverso i muri *sia* afferrare un asciugamano che cade? Come è possibile che la sostanza mentale possa *sia* eludere tutti i rilevamenti fisici *sia* controllare il corpo? Uno spettro nella macchina non ci è di nessun aiuto nei nostri sforzi teorici a meno che non sia uno

2.3

spettro che possa spostare le cose – come un rumoroso folletto che può rovesciare una lampada o sbattere una porta – ma qualsiasi cosa in grado di muovere un oggetto fisico è essa stessa un oggetto fisico (benché forse un tipo strano e finora non studiato di oggetto fisico).

Che dire, allora, della possibilità di sostenere che la sostanza mentale sia effettivamente un tipo speciale di materia? Nelle sedute spiritiche dell'epoca vittoriana, i medium creavano dal nulla qualcosa che chiamavano «ectoplasma», una strana sostanza appiccicosa che veniva considerata la materia basilare del mondo spirituale, ma che poteva essere intrappolata in barattoli di vetro e che poteva inumidire, colare e riflettere la luce come qualunque altra materia. Queste fraudolente messe in scena non ci dovrebbero dissuadere dal chiederci, più sobriamente, se la sostanza mentale non possa essere in effetti un qualcosa al di sopra e al di là degli atomi e delle molecole che compongono il cervello, ma pur sempre un tipo di materia scientificamente indagabile. L'ontologia di una teoria è il catalogo delle cose e dei tipi di cose che la teoria ritiene esistenti. L'ontologia delle scienze fisiche una volta includeva «il calorico» (la sostanza di cui era fatto il calore) e «l'etere» (la

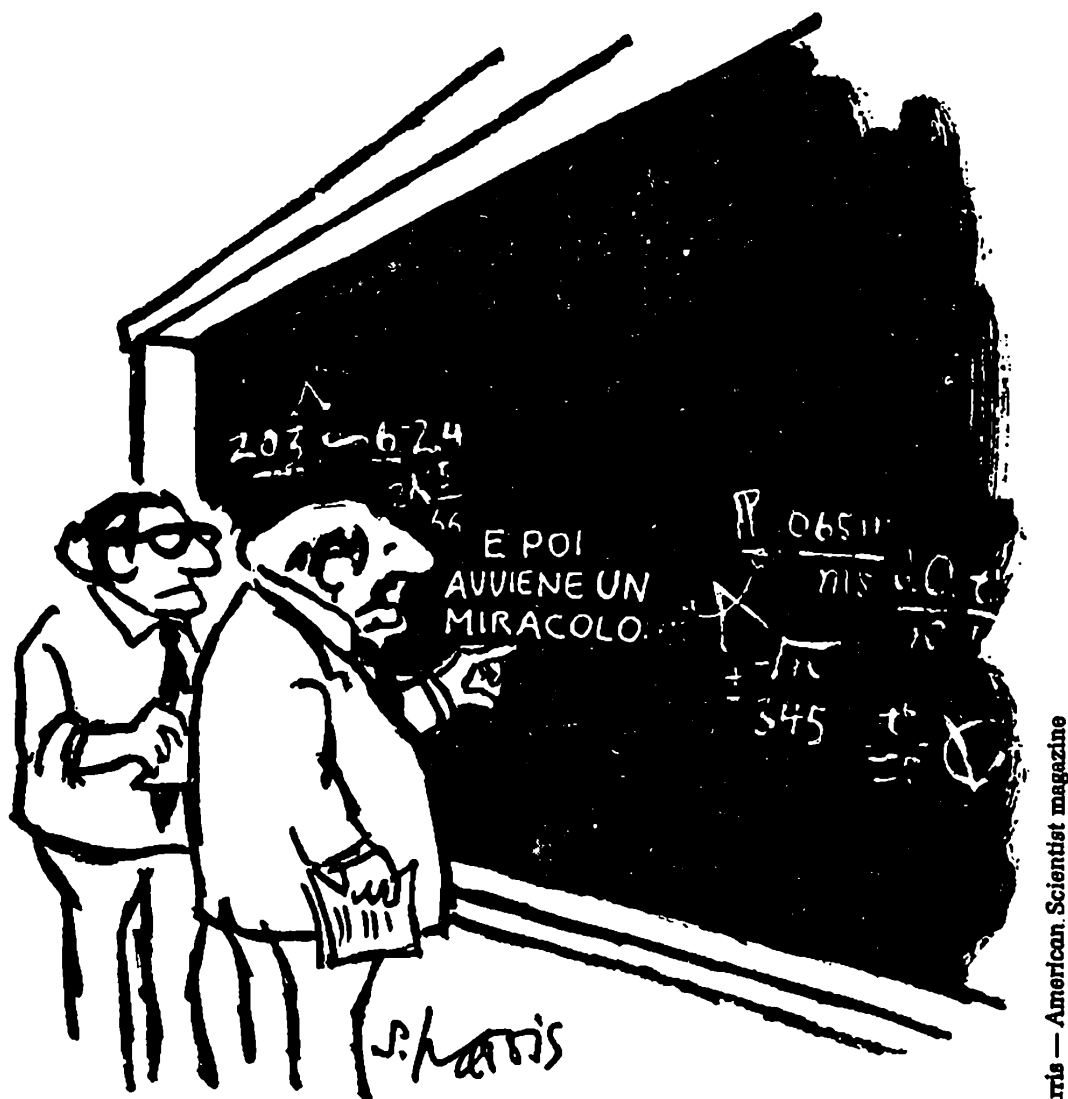
sostanza che riempiva lo spazio e che era il mezzo in cui si propagavano le vibrazioni luminose, così come l'aria o l'acqua sono il mezzo per le vibrazioni acustiche). Nessuno prende più seriamente queste cose, mentre neutrini, antimateria e buchi neri sono oggi inclusi nell'usuale ontologia scientifica. Forse è necessario un allargamento significativo dell'ontologia delle scienze fisiche per riuscire a spiegare i fenomeni della coscienza.

Una simile rivoluzione della fisica è stata di recente proposta dal fisico e matematico Roger Penrose, nel suo *La mente nuova dell'imperatore* (1989). Benché io non ritenga che egli sia riuscito a giustificare la sua rivoluzione,[3] è importante notare come egli sia stato attento a *non* cadere nella trappola del dualismo. Cosa intendo sottolineare? Penrose dice esplicitamente che con la sua rivoluzione egli intende rendere la mente cosciente *più* accessibile all'indagine scientifica, e non meno. Sicuramente non è un caso che i pochi dualisti che apertamente confessano le proprie opinioni abbiano tutti ammesso candidamente e tranquillamente che non hanno alcuna teoria di sorta che spieghi come lavora la mente – qualcosa, insistono, che è al di là della comprensione umana.[4] C'è il vago sospetto che la caratteristica più attraente della sostanza mentale sia la sua promessa di essere *così* misteriosa da tenere la scienza in scacco per sempre.

Questo atteggiamento profondamente antiscientifico del dualismo è, a mio avviso, la sua caratteristica più squalificante ed è il motivo che spiega perché in questo libro io abbia adottato la regola apparentemente dogmatica di evitare il dualismo *a tutti i costi*. Non che io pensi di riuscire a dare una di quelle dimostrazioni che stendono al tappeto la tesi avversaria, che proclami il dualismo, in tutte le sue forme, falso e incoerente, ma, visto il modo in cui il dualismo sguazza nel mistero, *accettare il dualismo significa rinunciare a capire.*

Su questo c'è un accordo generale, ma è tanto superficiale quanto diffuso, e nasconde qualche crepa preoccupante nel muro materialista. Scienziati e filosofi possono aver raggiunto un certo consenso in favore del materialismo, ma, come vedremo, sbarazzarci delle vecchie prospettive dualiste è più difficile di quanto abbiano pensato i materialisti contemporanei. Trovare i sostituti adatti per le tradizionali concezioni dualiste richiederà alcune rettifiche piuttosto sorprendenti del nostro modo abituale di pensare, rettifiche che sembreranno dapprincipio sicuramente controintuitive sia per gli scienziati sia per i profani.

Il fatto che la mia teoria sembri dapprincipio in forte

2.4 "CREDO CHE DOVRESTI ESSERE UN PO' PIÙ ESPLICITO QUI NEL PUNTO DUE".

© 1975 Sidney Harris — American Scientist magazine

conflitto con le concezioni diffuse, non lo considero di cattivo augurio. Al contrario, sarebbe errato aspettarsi che una buona teoria della coscienza costituisca una facile lettura – una di quelle che immediatamente ci fanno risuonare un campanello familiare, che ci fanno esclamare, con un malcelato orgoglio: «Certo! Lo sapevo già! È ovvio, ora che mi ci fai pensare!». Se da qualche parte ci fosse una teoria del genere, ormai l'avremmo sicuramente trovata. Ci balocchiamo con i misteri della mente da così lungo tempo e con così scarsi risultati che dobbiamo prendere sul serio la possibilità che alcune cose su cui ci troviamo tutti d'accordo non sono poi così ovvie come sembrano. I miei candidati verranno presto presentati.

Oggi alcuni ricercatori sul cervello – forse la parte più ampia e flemmatica di loro – continuano a far finta che il cervello sia, per loro, soltanto un altro organo, come i reni e il pancreas, e che deve essere descritto e spiegato solo nei termini sicuri delle scienze fisiche e biologiche. Non si sognerebbero mai di menzionare la mente o qualcosa di «mentale» nel corso dei loro impegni professionali. Per altri ricercatori, teoreticamente più audaci, esiste un nuovo oggetto di studio: il cervello/mente (Churchland, 1986). Questo fortunato neologismo esprime chiaramente il dominante materialismo di tali ricercatori, che apertamente ammettono che ciò che rende il cervello particolarmente affascinante e sconcertante è il fatto che in un modo o nell'altro *è* la mente. Ma anche tra questi ricercatori c'è una certa riluttanza a confrontarsi con i Grandi Argomenti, un desiderio di posporre indefinitamente le imbarazzanti domande sulla natura della coscienza.

Ma sebbene questo atteggiamento sia del tutto ragionevole – non voglio misconoscere il valore della strategia del *divide et impera* – ha l'effetto di distorcere alcuni dei nuovi concetti che sono sorti nel campo di quella che oggi viene chiamata *scienza cognitiva*. Quasi tutti i ricercatori della scienza cognitiva, che si considerino neuroscienziati o psicologi o ricercatori dell'intelligenza artificiale, tendono a posporre le questioni sulla coscienza restringendo la loro attenzione ai sistemi «periferici» e «subordinati» della mente/cervello, ai quali viene attribuito il compito di rifornire e servire un «centro» vagamente immaginato dove hanno luogo «il pensiero cosciente» e «l'esperienza». Ne risulta una tendenza a lasciare che «il centro» si occupi di una parte troppo grande del lavoro della mente e ciò conduce i ricercatori a sottovalutare la «quantità di comprensione» che può essere svolta dai sistemi relativamente periferici del cervello (Dennett, 1984b).

Per esempio è una tendenza molto diffusa tra i ricercatori quella di pensare che i sistemi percettivi forniscano «segnali in ingresso» a qualche arena pensante centrale, che a sua volta fornisce «controllo» o «direzione» a qualche sistema relativamente periferico che governa i movimenti corporei. Si immagina inoltre che tale arena centrale possa servirsi del materiale grezzo immagazzinato in vari sistemi della memoria più o meno subordinati. Ma la stessa idea che ci possa essere una qualche importante divisione teoretica tra tali presunti sottosistemi come «memoria a lungo termine» e «ragionamento» (o «progettazione») è più un prodotto della strategia del *divide et impera* che

qualcosa che rispecchi la natura. Come vedremo presto, l'attenzione esclusiva verso gli specifici sottosistemi del cervello/mente causa spesso una specie di miopia teoretica che impedisce ai ricercatori di vedere che i loro modelli presuppongono ancora che da qualche parte, convenientemente nascosto nell'oscuro «centro» del cervello/mente, ci sia un Teatro Cartesiano, un luogo in cui «tutto converge» e la coscienza si verifica. Questa può sembrare una buona idea, un'idea inevitabile, ma fino a che non vedremo, un po' più dettagliatamente, perché non lo è, il Teatro Cartesiano continuerà ad attrarre miriadi di ricercatori paralizzati da un'illusione.

5. *La sfida*

Nel paragrafo precedente ho affermato che se non possiamo avere nulla di meglio del dualismo, allora non possiamo comprendere la coscienza umana. Alcune persone sono convinte che ciò ci è impossibile in ogni caso. Questo disfattismo, oggi, mentre siamo sommersi da un profluvio di progressi scientifici che aspettano di essere sfruttati, mi sembra ridicolo e perfino patetico, ma riconosco che potrebbe essere la triste verità. Forse la coscienza non può realmente essere spiegata, ma come potremmo esserne certi se nessuno ci prova? Io penso che molti – effettivamente, la maggior parte – dei pezzi del puzzle sono già ben compresi e a me non rimane che adoperarmi per farli combaciare. Quelli che vogliono difendere la Mente contro la Scienza dovrebbero augurarmi buona fortuna in questo tentativo, perché, se hanno ragione, il mio progetto fallirà in ogni caso, ma, se io svolgo il mio lavoro nel migliore dei modi, il mio fallimento getterà della luce sul perché la scienza sia condannata all'impotenza. Almeno riusciranno ad avere un argomento contro la scienza, ed io avrò fatto tutto il lavoro spiacevole per loro.

Le regole fondamentali del mio progetto sono chiare:

1. *Non è permesso nessun tessuto miracoloso.* Tenterò di spiegare ogni caratteristica spinosa della coscienza umana rimanendo all'interno del quadro teorico della scienza fisica contemporanea; in nessun punto farò appello a forze, sostanze o poteri organici inesplicabili e sconosciuti. In altre parole, mi dichiaro conservatore: intendo indagare che cosa si può fare senza uscire dai limiti attuali della scienza, riservandomi di

invocare una rivoluzione nel materialismo solo come ultima risorsa.

2. *Nessuna simulazione di anestesia.* Dei comportamentisti è stato detto che fingono di essere sotto anestesia – pretendono di non avere quelle esperienze che noi sappiamo benissimo che essi hanno come tutti noi. Se desidero negare l'esistenza di qualche caratteristica controversa della coscienza, spetta a me *mostrare* che è in qualche modo illusoria.

3. *Nessuna pignoleria sui dettagli empirici.* Cercherò di presentare correttamente tutti i fatti scientifici oggi conosciuti, ma non è affatto scontato quali siano i progressi importanti che resisteranno alla verifica del tempo. Se dovessi limitarmi ai «fatti che si sono conquistati un posto nei libri di testo», mi precluderei l'utilizzazione di alcune delle scoperte più recenti e più illuminanti (se questo è ciò che sono). E finirei *in ogni caso* per fornire involontariamente alcune falsità, se la storia recente ci insegna qualcosa. Alcune delle «scoperte» sulla visione per le quali David Hubel e Torstein Wiesel hanno meritatamente ricevuto il premio Nobel nel 1981 cominciano ora a sfilacciarsi, e la famosa teoria «retinex» della visione cromatica di Edwin Land, che è stata considerata dalla maggior parte dei filosofi della mente e da altri non specialisti come un fatto assodato per più di venti anni, non è più così in auge tra gli scienziati della visione.[5]

Quindi, dal momento che come filosofo sono interessato a garantire delle *possibilità* (e a refutare le dichiarazioni d'impossibilità), io proporrò degli abbozzi di teoria piuttosto che teorie completamente sviluppate e empiricamente confermate. Un abbozzo di teoria, o un modello di come il cervello *potrebbe* fare qualcosa, può trasformare una situazione confusa in un programma di ricerca: sebbene questo modello non funzioni a pieno, esiste qualche altra variante più realistica in grado di portare a termine la faccenda? (L'abbozzo di spiegazione, nel capitolo 1, del come vengono prodotte le allucinazioni ne è un esempio.) Un tale abbozzo è direttamente e esplicitamente vulnerabile da una refutazione empirica, ma, se si vuole sostenere che il mio abbozzo non è una spiegazione *possibile* del fenomeno, bisogna far vedere che cosa *è obbligato* a lasciar fuori o cosa *non può* fare; se ci si limita ad affermare che il mio modello può essere incoerente in molti dettagli, potrò essere d'accordo.

Ciò che non va con il dualismo cartesiano, per esempio, non è il fatto che Cartesio abbia scelto la ghiandola pineale – e non il talamo o, diciamo, l'amigdala – come sito dell'interazione con la mente, ma è *l'idea stessa* di un tale sito di interazione tra la mente e il cervello. Ciò che viene considerato «pignoleria» muta, naturalmente, man mano che la scienza avanza e, inoltre, differenti scienziati hanno standard differenti. Io tenterò di errare per eccesso di specificità, non solo per amplificare le differenze con la tradizionale filosofia della mente, ma anche per dare ai critici del versante empirico un bersaglio più chiaro a cui sparare.

In questo capitolo abbiamo incontrato le caratteristiche basilari del mistero della coscienza. La misteriosità stessa della coscienza è una delle sue caratteristiche centrali – forse è perfino una sua caratteristica vitale, senza la quale non può sopravvivere. E poiché questa possibilità, sebbene molto confusamente, è ampiamente considerata, la prudenza tende a favorire quelle dottrine che non pretendono neanche di spiegare la coscienza, giacché la coscienza ha un'enorme importanza per noi. Il dualismo, l'idea che un cervello non possa essere una cosa pensante di modo che una cosa pensante non possa essere un cervello, è attraente per svariate ragioni, ma dobbiamo resistere alla tentazione: adottare il dualismo equivale davvero ad accettare la sconfitta senza ammetterlo. Adottare il materialismo non serve di per sé a dissolvere gli enigmi sulla coscienza, e non serve neppure una serie di inferenze dirette ricavate dalla scienza del cervello. In qualche modo il cervello deve essere la mente, ma, fintanto che non riusciamo a capire un po' più dettagliatamente come ciò sia possibile, il nostro materialismo non spiegherà la coscienza, ma prometterà soltanto di riuscire a spiegarla in un futuro radioso. Questa promessa non può essere mantenuta, ho già accennato, fino a che non impareremo ad abbandonare una porzione più ampia dell'eredità cartesiana. Allo stesso tempo, per quanto forte possa essere la capacità esplicativa delle teorie materialiste, esse non avranno spiegato la coscienza se trascureranno quei fatti sull'esperienza che noi conosciamo così intimamente «dall'interno». Nel prossimo capitolo, stenderemo l'inventario iniziale di questi fatti.

3

UNA VISITA AL GIARDINO FENOMENOLOGICO

1. *Benvenuti a fenom*

Supponiamo che un folle sostenga che non esistono gli animali. Potremmo decidere di metterlo di fronte al suo errore portandolo allo zoo e dicendogli: «Guarda! Che cosa sono queste cose, se non animali?». Non ci aspetteremmo con ciò di guarirlo, ma almeno ci saremmo presi la soddisfazione di mettere in chiaro con noi stessi l'enormità della pazzia che ci è toccato sentire. Ma supponiamo che egli dica allora: «Oh, lo so perfettamente che esistono queste cose – leoni, struzzi, boa – ma che cosa ti fa pensare che questi cosiddetti animali siano effettivamente *animali*? In realtà, non sono altro che robot coperti di pelliccia – beh, veramente, alcuni sono ricoperti di piume e altri di scaglie». Questa potrebbe ancora essere una pazzia, ma un tipo differente e più difendibile di pazzia. Questo folle potrebbe semplicemente avere un'idea rivoluzionaria sulla natura ultima degli animali.[1]

Gli zoologi sono gli esperti sulla natura ultima degli animali, e i giardini zoologici – gli zoo – servono all'utile scopo educativo di far conoscere alla gente l'argomento di loro competenza. Se gli zoologi dovessero scoprire che il nostro folle aveva ragione (in qualche senso lato), si accorgerebbero che gli zoo sarebbero utilissimi per tentare di spiegare la loro scoperta. Potrebbero dire: «Abbiamo scoperto che gli animali – quelle cose cioè che tutti abbiamo visto allo zoo – non sono esattamente come pensavamo che fossero; sono, in effetti, così differenti che non dovremmo chiamarli propriamente animali. Quindi, capite bene che in realtà *non esiste* alcun animale nel vero senso della parola».

I filosofi e gli psicologi spesso usano il termine *fenomenologia* come un termine generale che si riferisce a tutti gli oggetti – la flora e la fauna, si potrebbe dire – che popolano la nostra

esperienza cosciente: pensieri, odori, pruriti, dolori, immaginarie mucche viola, sospetti e altro. Ci sono vari antecedenti piuttosto ben definiti di un tale uso che vale la pena ricordare. Nel XVIII secolo, Kant distinse i *fenomeni*, le cose come appaiono, dai *noumeni*, le cose in sé; durante lo sviluppo delle scienze naturali e fisiche nel XIX secolo, il termine *fenomenologia* venne usato per riferirsi allo studio meramente descrittivo di qualunque argomento, in modo neutrale e pre-teoretico. La fenomenologia del magnetismo, per esempio, era ben conosciuta da William Gilbert nel XVI secolo, ma la sua spiegazione dovette attendere fino al secolo scorso quando si chiarirono le relazioni tra magnetismo e elettricità, e furono conosciuti i lavori teorici di Faraday, Maxwell e altri. Accentuando questa divisione tra l'acuta osservazione e la spiegazione teoretica, all'inizio di questo secolo si sviluppò, attorno alla figura di Edmund Husserl, la scuola o il movimento filosofico noto sotto il nome di Fenomenologia (con la F maiuscola). Il suo scopo era quello di trovare una nuova fondazione per tutta la filosofia (in effetti, per tutta la conoscenza) basandosi su una tecnica speciale di introspezione in cui il mondo esterno e tutte le sue implicazioni e presupposizioni fossero «messi tra parentesi» grazie ad un particolare atto della mente chiamato *epoché*; da questa sottrazione doveva risultare uno stato indagatore della mente che avrebbe dovuto permettere al Fenomenologo di essere messo in contatto con i puri oggetti dell'esperienza cosciente, chiamati *noemata*. Come altri tentativi di spogliarsi dell'interpretazione e rivelare i fatti elementari della coscienza all'osservazione rigorosa, quali il movimento impressionistico nelle arti e le psicologie introspezioniste di Wundt, Tichener e altri, la Fenomenologia non è riuscita a trovare un unico metodo ben stabilito su cui tutti potessero essere d'accordo.

Così, sebbene esistano gli zoologi, non esistono in realtà i fenomenologi: esperti incontestati della natura delle cose che sguazzano nel flusso della coscienza. Ciononostante possiamo seguire la consuetudine recente e adottare il termine (con la f minuscola) come un termine generico per i vari oggetti nell'esperienza cosciente che devono essere spiegati.

Una volta, in un mio articolo intitolato *On the Absence of Phenomenology* (1979), tentai di sostenere qualcosa che assomiglia alla seconda parte della pazzia in questione: le cose di cui è composta la coscienza sono in realtà così differenti da come la gente le immagina che sarebbe bene non usare i vecchi termini. Ma questa è sembrata a qualcuno una proposta così esagerata

(«Come diamine si può pensare che ci sbagliamo sulla nostra vita interiore!») da essere accomunata al primo tipo di pazzia («Dennett pensa che non esistano dolori o aromi o sogni ad occhi aperti!»). Questa è una descrizione caricaturale della disputa, naturalmente, ma è molto attraente. Me la sarei potuta cavare meglio se avessi avuto a portata di mano un giardino fenomenologico – un fenom – da usare per le mie spiegazioni; avrei potuto dire: «Ci siamo accorti che le cose che sguazzano nel flusso della coscienza – quelle cose cioè come i dolori e gli aromi, i sogni ad occhi aperti e le immagini mentali, gli scatti di odio e di lussuria, e tutti i normali abitatori di fenom – non sono come pensavamo precedentemente che fossero. Sono in realtà così differenti che sarebbe bene trovare qualche nuovo termine per loro».

Facciamo quindi una piccola passeggiata nel nostro giardino fenomenologico, giusto per rassicurarci che sappiamo di cosa stiamo parlando (anche se non sappiamo *ancora* con certezza quale sia la natura ultima di queste cose). Sarà un giro introduttivo deliberatamente superficiale, al solo scopo di indicare gli oggetti, dire qualche parola informativa e sollevare alcune questioni, prima di immergerci nelle speculazioni più serie nel resto del libro. Poiché mi troverò presto costretto a presentare delle sfide colossali al pensiero comune, non vorrei che qualcuno pensasse che io semplicemente ignori tutte quelle cose meravigliose che abitano le menti delle *altre persone*.

Fenom si divide in tre parti: (1) *le esperienze del mondo «esterno»*, come le cose viste, udite, odorate, le sensazioni di scivoloso o ruvido, di caldo o freddo, e delle posizioni dei nostri arti; (2) *le esperienze del mondo puramente «interno»*, come le immagini della fantasia, i suoni e le visioni interne dei sogni ad occhi aperti e dei monologhi interiori, le memorie, le idee brillanti e le improvvise preoccupazioni; (3) *le esperienze delle emozioni o degli «affetti»* (per usare un termine inadeguato in voga tra gli psicologi), che vanno dai dolori corporei, il solletico e le «sensazioni» di fame e sete, attraverso stati emozionali intermedi come gli accessi di ira, gioia, odio, imbarazzo, lussuria, fino agli episodi più lontani dal corporeo come l'orgoglio, l'ansia, il rimpianto, il distacco ironico, il pentimento, il timore reverenziale, la calma glaciale.

Non pretendo alcuna accuratezza per questa divisione tripartita (esterno, interno e affetti); come un serraglio in cui si trovano i pipistrelli con gli uccelli e i delfini con i pesci, questa tassonomia è dovuta più alla superficiale similitudine e a una

dubbia tradizione che a qualche più profonda relazione tra i fenomeni; ma da qualche parte bisogna pur iniziare e qualunque tassonomia che ci fornisce una base ci aiuterà a non trascurare del tutto qualche specie.

2. *La nostra esperienza del mondo esterno*

Cominciamo dalle nostre modalità sensoriali più grossolane: il gusto e l'odorato. Come molta gente sa, le nostre papille gustative sono sensibili in realtà solo al dolce, l'acido, il salato e l'amaro, e per lo più noi «gustiamo con il naso», il che spiega perché i cibi perdono il loro sapore quando siamo raffreddati. L'epitelio nasale è per l'olfatto ciò che la retina è per la vista. Esiste una grande varietà di cellule epiteliali, ognuna sensibile a un tipo differente di molecola trasportata dall'aria. Ciò che è importante in definitiva è la *forma* delle molecole. Le molecole fluttuano nel naso e come tanti martelletti microscopici battono sulle specifiche cellule sensoriali nell'epitelio. Spesso è possibile percepire facilmente le molecole in concentrazioni sorprendentemente basse, anche di poche parti per miliardo. Altri animali hanno un olfatto enormemente più sviluppato del nostro, non solo sono capaci di discriminare più odori, in tracce più deboli (il segugio ne è un esempio familiare), ma hanno anche una migliore risoluzione temporale e spaziale degli odori. Il perché di questo fatto è chiaro: le molecole *vagano* più o meno casualmente fino ad arrivare nelle nostre cavità nasali e il loro arrivo in un punto specifico dell'epitelio fornisce scarse informazioni sul luogo esterno di provenienza; a differenza dei fotoni che viaggiano in linea retta attraverso la pupilla, arrivando a una localizzazione sulla retina che corrisponde geometricamente a una sorgente esterna o a un tragitto dalla sorgente esterna. Se la risoluzione della nostra vista fosse così debole come quella del nostro olfatto, quando un uccello passa sopra la nostra testa il cielo per un po' diventerebbe per noi *tutto uccelloso.* (Alcune specie hanno una vista così debole – cioè, con una risoluzione e discriminazione non migliore di questa – ma che cosa si prova, se mai si prova qualcosa, a vedere le cose in questa maniera è un altro problema, al quale ritorneremo in un capitolo successivo.)

I nostri sensi del gusto e dell'odorato sono accoppiati fenomenologicamente, così come lo sono quello tattile e quello cinestetico, il senso che ci indica la posizione e il movimento dei

nostri arti e delle altre parti del corpo. Noi «sentiamo» le cose toccandole, afferrandole, spingendole in varie maniere, ma le sensazioni coscienti che ne risultano, pur sembrando ad una riflessione superficiale le dirette «traduzioni» della stimolazione dei recettori tattili sotto la pelle, sono ancora una volta il prodotto di un elaborato processo di integrazione delle informazioni provenienti da svariate fonti. Bendati e prendi in mano un bastoncino (o una penna o una matita). Tocca varie cose attorno a te con questa bacchetta e nota come puoi distinguere facilmente la loro granulosità superficiale – come se il tuo sistema nervoso avesse dei sensori sulla punta della bacchetta. È necessario uno sforzo speciale, e spesso senza risultati, per cogliere il modo in cui il bastone preme sui tuoi polpastrelli, il modo in cui vibra o resiste al movimento quando incontra varie superfici. Queste transazioni tra bastone e ricettori tattili sotto la pelle (aiutate la maggior parte delle volte da suoni poco notati) forniscono le informazioni che il tuo cervello integra in un riconoscimento cosciente della granulosità superficiale della carta, del cartone, della lana o del vetro, ma questi complicati processi d'integrazione sono quasi trasparenti per la coscienza. Cioè, non sappiamo – non possiamo – dire come «noi» lo facciamo. Se vuoi un esempio ancora più indiretto, pensa a come puoi sentire la scivolosità di una macchia d'olio sulla strada sotto le ruote della tua automobile mentre fai una curva. Il punto focale fenomenologico del contatto è il punto in cui le gomme toccano l'asfalto e non un punto qualsiasi del tuo corpo innervato, che sta seduto e vestito sul sedile dell'automobile o delle tue mani inguantate sullo sterzo.

Ora, rimanendo sempre bendato posa la bacchetta e fatti porgere da qualcuno un pezzo di ceramica, uno di plastica, uno di legno levigato e uno di metallo. Sono tutti estremamente lisci e scivolosi, eppure non avrai alcuna difficoltà a riconoscere la loro specifica levigatezza – e non certo perché hai dei ricettori specializzati per la plastica o la ceramica nei tuoi polpastrelli. La differenza nella conducibilità termica delle sostanze è a prima vista il fattore più importante, ma non è essenziale: ti potrà sorprendere la facilità con cui puoi distinguere talvolta le varie superfici «sentendole» con la bacchetta. Questi successi devono dipendere dalle vibrazioni che senti raccolte nella bacchetta o dalle differenze indescrivibili – ma percepibili – nei rumori di graffiatura o battitura. Ma *sembra* proprio che alcune delle tue terminazioni nervose siano nella bacchetta, giacché tu *senti* le differenze delle superfici *nella* punta della bacchetta.

Consideriamo, ora, l'udito. La fenomenologia dell'udito consiste di tutti i tipi di suoni che possiamo ascoltare: la musica, le parole pronunciate, i boati e i fischi, i cinguettii e gli scoppi. I pensatori che riflettono sull'udito sono spesso tentati di «intonare la piccola banda nella testa». Questo è un errore, e, per essere sicuri di individuarlo ed evitarlo, voglio renderlo più vivido con l'aiuto di una favola.

C'era una volta, circa a metà del secolo scorso, un inventore pazzo impegnato in un dibattito con un filosofo con i piedi per terra, Phil. L'inventore aveva annunciato che il suo obiettivo era quello di costruire un congegno che potesse «registrare» automaticamente per poi «riprodurre» ad «alta fedeltà» un'orchestra e un coro che eseguivano la Nona Sinfonia di Beethoven. Un'assurdità, disse Phil. È assolutamente impossibile. Posso immaginare facilmente un congegno meccanico che registra la sequenza in cui i tasti del pianoforte vengono pigiati e che poi governa la riproduzione di tale sequenza su un piano appositamente preparato – si potrebbe farlo utilizzando un rotolo di carta perforata, per esempio – ma pensa all'enorme varietà di suoni e dei loro modi di produzione in un'interpretazione della Nona di Beethoven! Ci sono un centinaio di voci umane di estensione e timbro differenti, dozzine di strumenti ad arco, di ottoni, fiati e percussioni. Il congegno che dovrebbe riprodurre insieme una tale varietà di suoni sarebbe un'ingombrante mostruosità che farebbe impallidire il più potente organo da chiesa – e, se suonasse con «l'alta fedeltà» che proponi, dovrebbe senza dubbio incorporare, quasi letteralmente, un drappello di schiavi umani per eseguire le parti vocali e ciò che tu chiami la «registrazione» di una particolare esecuzione con tutte le sue sfumature dovrebbe essere composta da centinaia di spartiti – uno per ogni musicista – con migliaia o anche milioni di annotazioni.

L'argomento di Phil è ancora stranamente potente; è sorprendente che tutti quei suoni possano essere fedelmente codificati grazie a una trasformata di Fourier in un'unica linea ondulata incisa su un disco di resina vinilica o rappresentata magneticamente su un nastro. È ancora più sorprendente che un unico cono di carta, messo in vibrazione da un elettromagnete guidato da un'unica linea ondulata, possa fare giustizia sia a uno squillo di tromba, che a uno strimpellio di banjo, a un discorso umano, al rumore di una bottiglia piena di vino che si infrange sul marciapiede. Phil non poteva

immaginare nulla di così potente e ha scambiato la sua mancanza d'immaginazione per una profonda intuizione sulla necessità.

La «magia» delle trasformate di Fourier ci apre un nuovo campo di possibilità a cui pensare, ma dovremmo notare che non elimina di per sé il problema che ha confuso Phil; lo pospone meramente. Giacché se dall'alto del nostro progresso possiamo sorridere di Phil che non è riuscito a capire come la sequenza della rarefazione e compressione dell'aria che stimola l'orecchio possa essere registrata e riprodotta, ci accorgeremo che tale sorrisetto scomparirà dalla nostre facce non appena riflettiamo alla domanda *successiva*: che cosa accade ai segnali una volta che l'orecchio li ha correttamente ricevuti?

Dall'orecchio un'ulteriore batteria codificata di sequenze di segnali modulati (ma ora parzialmente analizzati e suddivisi in flussi paralleli, che ricordano malauguratamente le centinaia di partiture di Phil) viaggia verso l'interno, verso la parte più oscura del cervello. Queste sequenze di segnali, come le linee ondulate del disco, non sono ovviamente dei *suoni sentiti*, sono successioni di impulsi elettrochimici che risalgono gli assoni dei neuroni. Non dovrebbe esserci un posto ancora più centrale nel cervello dove queste sequenze di segnali controllano la performance teatrale di un portentoso organo della mente? Quando avviene, dopo tutto, la *traduzione* finale di questi segnali senza tono in suoni soggettivamente uditi?

Non vogliamo cercare dei posti nel cervello che vibrano come corde di chitarra, più di quanto vogliamo trovare dei posti nel cervello che diventano viola quando immaginiamo una mucca viola. Questi sono chiaramente dei vicoli ciechi, ciò che Gilbert Ryle (1949) chiamerebbe «errori categoriali». Ma allora che cosa *potremmo* trovare nel cervello che possa convincerci di aver raggiunto la fine del viaggio dell'esperienza auditiva?[2] Come può un qualunque complesso di proprietà fisiche di eventi nel cervello equivalere alle – o essere anche solamente la spiegazione delle – emozionanti proprietà dei suoni che udiamo?

Dapprincipio queste proprietà non sembrano analizzabili – o, per usare un termine caro ai fenomenologi, sembrano *ineffabili*. Ma è possibile fare in modo che almeno alcune di queste proprietà apparentemente atomiche e omogenee vengano scomposte e descritte. Prendi una chitarra e pizzica la corda più grande, quella del mi più basso, senza pigiare alcun tasto. Ascolta attentamente il suono. È composto di parti in qualche

modo descrivibili o sembra un tutt'uno, inscindibile e ineffabile, qualcosa di inconfondibilmente «chitarrico»? Molti opteranno per la seconda maniera di descrivere la loro fenomenologia. Ora pizzica di nuovo la corda libera e poi con il dito premi attentamente e leggermente l'ottavo tasto per creare un «armonico» alto. Improvvisamente senti un nuovo suono: in qualche modo «più puro» e naturalmente un'ottava più alto. Alcune persone insistono nel dire che questo è un suono affatto nuovo, mentre altri descrivono l'esperienza dicendo che «il basso è caduto via dalla nota» lasciando solo l'acuto. Poi pizzica la corda libera una terza volta. Questa volta puoi sentire, con sorprendente distinzione, il tono armonico superiore che è stato isolato con la seconda pizzicata. La omogeneità e l'ineffabilità della prima esperienza è scomparsa, sostituita da una dualità così direttamente rilevabile e chiaramente descrivibile come quella tra due corde.

La differenza nell'esperienza è sbalorditiva, ma la nuova complessità appresa nella terza pizzicata *era sempre stata lì* (pronta ad essere discriminata). La ricerca ha mostrato che solo grazie alla complessa tavola degli armonici siamo capaci di riconoscere il suono di una chitarra da quello di un liuto o di un clavicembalo. Tali ricerche possono aiutarci a *spiegare* le differenti proprietà delle esperienze auditive, analizzando i componenti informativi e i processi che le integrano, permettendoci così di prevedere e anche provocare sinteticamente specifiche esperienze auditive, ma sembra tuttavia che lascino intatto il problema di dire *a cosa equivalgono* tali proprietà. Perché mai gli armonici prodotti dalla chitarra sembrano *così* e quelli del liuto sembrano *cosà*? Non abbiamo ancora dato una risposta a questo interrogativo residuo, anche se l'abbiamo alleggerito facendo vedere che almeno alcune delle proprietà che sembravano inizialmente ineffabili sono dopo tutto passibili di una certa analisi e descrizione.[3]

Gli studi che si occupano dell'elaborazione delle percezioni auditive ci invitano a pensare che esistano dei meccanismi specializzati per decifrare le diverse specie di suoni, simili in un certo senso ai componenti immaginati da Phil nella sua fantastica macchina riproduttrice dei suoni. I suoni di un discorso, in particolare, sembrano gestiti da meccanismi appositi – come direbbe un ingegnere. La fenomenologia della percezione dei discorsi ci suggerisce che in qualche parte del cervello avvenga una generale ricostruzione dei segnali in ingresso, che ci sia nel cervello una sorta di congegno analogo a quello di uno studio di

registrazione musicale, dove molteplici canali di registrazione vengono missati e aggiustati in vario modo per creare un «master» stereofonico da cui le successive registrazioni vengono copiate sui vari supporti.

Per esempio, nei discorsi nella nostra lingua madre, noi percepiamo le singole parole separate da leggere pause silenziose; abbiamo, cioè, la chiara sensazione del confine tra le parole, che non potendo essere contraddistinto da un diverso colore o da lineette, né segnalato da uno squillo o da uno scampanellio, non può essere indicato che da un intervallo silenzioso di varia durata – come le pause che separano le lettere e le parole nell'alfabeto Morse. Se, in un esperimento, si chiede a un soggetto di prestare attenzione alle pause tra le parole, egli non ha la minima difficoltà a individuarle. Sembra che queste pause siano lì. Ma se osserviamo il profilo dell'energia acustica del segnale in ingresso, noteremo che le zone a più bassa energia (i momenti più vicini al silenzio) non combaciano affatto con i confini tra le parole. La segmentazione dei suoni linguistici è un processo che impone dei confini basati sulla struttura grammaticale del linguaggio e non sulla struttura fisica delle onde sonore (Liberman e Studdert-Kennedy, 1977). Questo ci aiuta a capire perché i discorsi in una lingua straniera ci sembrano una sequela impetuosa e ingarbugliata di suoni non segmentati: lo «studio di registrazione», l'apposito meccanismo nel cervello, è privo della necessaria cornice grammaticale per individuare i corretti segmenti, così la cosa migliore che può fare è semplicemente quella di passare i segnali in arrivo così come sono, senza alcun sostanziale ritocco.

Quando percepiamo un discorso noi non siamo consapevoli solo dell'individualità e delle categorie grammaticali delle parole. (Se non fossimo consapevoli d'altro, non potremmo essere in grado di dire se stiamo ascoltando o leggendo le parole.) Le parole sono chiaramente demarcate, ordinate e identificate, ma ci si presentano anche vestite di proprietà più carnose. Per esempio, ho appena ascoltato la voce tipicamente britannica del mio amico Nick Humphrey, una voce cortesemente provocatoria, non proprio beffarda; mi sembra di *aver ascoltato* il suo sorriso e nella mia esperienza è presente la sensazione che dietro le parole ci fosse una risata, pronta ad esplodere come il sole dietro a qualche nuvola fuggevole. Noi siamo consapevoli di molte proprietà dei suoni di un discorso, non solo se l'intonazione è crescente o calante, ma anche se la voce è stridente, ansimante o pastosa, vi sappiamo riconoscere

perfino la lamentosità umorale, il tremolio della paura e la piattezza della depressione. E come abbiamo appena osservato nel caso della chitarra, quelle che sembrano dapprincipio delle proprietà del tutto atomiche e omogenee spesso sono suscettibili di analisi con un po' di pazienza e di pratica. Noi tutti riconosciamo il suono interrogativo di una domanda – e anche se ci sono o no sfumature dialettali in esso – ma è necessaria una certa sperimentazione sul tema e le sue variazioni, prima di poter descrivere con una certa accuratezza e sicurezza le differenze nelle forme di intonazione che conducono a questi differenti sapori auditivi.

In questo caso, «sapori» sembra proprio la metafora appropriata, senza dubbio perché le nostre capacità di analizzare i sapori sono così limitate. Le ben conosciute ma sempre sorprendenti dimostrazioni che noi gustiamo con il naso mostrano che i poteri del nostro gusto e del nostro olfatto sono così grossolani che abbiamo difficoltà ad identificare perfino la modalità tramite la quale siamo informati. Questa manchevolezza non è limitata al gusto e all'olfatto; noi ascoltiamo i toni di frequenza molto bassa – come le note più basse di un organo da chiesa – più con la sensazione delle vibrazioni nel nostro corpo che con le vibrazioni colte dai nostri orecchi. È sorprendente apprendere che «quel particolare Fa#, esattamente due ottave sotto il più basso Fa# che posso intonare», può essere *udito*, in realtà, con il mio fondo schiena piuttosto che con le mie orecchie.

Infine, rivolgiamoci brevemente alla vista. Quando i nostri occhi sono aperti noi abbiamo la sensazione di un ampio campo – spesso chiamato campo fenomenico o campo visivo – in cui le cose appaiono, con i loro colori, in quiete o in movimento, a una profondità o distanza variabile. Ingenuamente consideriamo quasi tutte le caratteristiche che rientrano nella nostra esperienza come proprietà oggettive delle cose esterne, osservate «direttamente» da noi, ma sin da bambini impariamo a riconoscere una categoria intermedia di oggetti – abbagli, luccichii, bagliori, macchie di colore – che sappiamo essere prodotti in qualche modo dall'interazione tra gli oggetti, la luce e il nostro apparato visivo. Noi continuiamo a considerare queste cose come «lì fuori» piuttosto che in noi, ma con alcune eccezioni: il fastidio che proviamo guardando il sole o una luce brillante subito dopo un periodo di adattamento all'oscurità, il nauseante roteare del campo fenomenico quando abbiamo le vertigini. Ci sembra più appropriato descrivere queste cose come «sensazioni negli occhi», più simili alle pressioni e ai pruriti che

sentiamo quando ci strofiniamo gli occhi piuttosto che a normali proprietà che si trovano lì fuori nelle cose che vediamo.

Tra le cose che vediamo lì fuori nel mondo fisico ci sono i quadri, le fotografie e tutte le immagini prodotte dagli uomini. Poiché tali cose ci appaiono soprattutto come «cose da vedere», noi siamo portati a dimenticare che esse costituiscono una recente acquisizione del nostro ambiente visivo, non più antica di poche decine di migliaia di anni. Grazie all'arte e alla tecnologia umana, noi siamo ora circondati di quadri, foto, mappe, disegni, sia immobili che in movimento. Queste immagini fisiche, che non sono che un tipo di «materiale grezzo» per l'elaborazione della percezione visiva, sono diventate un modello quasi irresistibile del «prodotto finito» della percezione visiva: «fotografie nella testa». Noi abbiamo la tendenza a dire: «Ovviamente, il risultato della *visione* è una fotografia nella testa (o nella mente). Cos'altro potrebbe essere? Non certo un suono o un sapore!» Ci occuperemo di questa curiosa ma diffusa malattia dell'immaginazione in molte maniere prima di aver finito, ma possiamo intanto notare che come le gallerie di fotografia per i ciechi sono uno sciupio di risorse, così le fotografie nella testa avranno bisogno di occhi nella testa per apprezzarle (per non parlare di un'appropriata illuminazione). Ma supponiamo pure che ci siano occhi mentali nella testa per apprezzare queste foto nella testa: che diremo, a loro volta, delle fotografie nella testa della testa prodotte da questi occhi interni? Come si potrà evitare un regresso all'infinito di fotografie e spettatori? Possiamo interrompere questo regresso solo scoprendo qualche spettatore che abbia una percezione che non crei un'altra fotografia che a sua volta necessiti di un altro spettatore. Forse il passo più indicato per interrompere questo regresso è proprio il primo?

Fortunatamente, ci sono altre ragioni che ci spingono ad essere scettici nei confronti di questa concezione della visione. Se la visione implicasse delle fotografie nella testa con cui noi (i nostri sé interiori) fossimo in un rapporto così intimo e speciale, fare dei disegni non dovrebbe essere più facile? Ricordati di quanto è difficile disegnare una rosa, per esempio. C'è la rosa, a grandezza naturale, a circa mezzo metro di fronte a te – sulla sinistra, diciamo, del tuo blocco da disegno. (Voglio davvero che immagini tutto questo con attenzione.) Sembra che tutti i dettagli visibili della rosa reale siano vividi, netti e facilmente accessibili, eppure il processo apparentemente semplice di spostare tutti quei dettagli di pochi gradi sulla destra in una copia

bidimensionale, in bianco e nero, è così impegnativo che la maggior parte della gente rinuncia e decide che semplicemente non sa disegnare. La traslazione delle tre dimensioni nelle due dimensioni è particolarmente difficile per noi, il che è un po' sorprendente perché la traslazione a prima vista inversa – vedere un'immagine realistica in due dimensioni come un oggetto o una situazione tridimensionale – è agevole e involontaria. Difatti, è proprio la difficoltà che incontriamo nel sopprimere questa interpretazione inversa che ci rende così gravoso anche il semplice compito di copiare uno schizzo.

Questo non è soltanto un problema di «coordinazione mano-occhio», perché anche le persone in grado di ricamare o assemblare orologi da polso con agevole destrezza possono rivelarsi completamente incapaci di copiare un disegno. Si potrebbe dire che è più un problema di coordinazione occhio-*cervello*. Quelli che padroneggiano l'arte sanno che richiede attenzioni particolari, trucchi come quello di non mettere perfettamente a fuoco ciò che si vede per facilitare il compito di sopprimere in qualche modo il contributo di *ciò che si sa* (la moneta è circolare, la tavola è rettangolare) affinché si possano osservare gli angoli reali sottesi alle linee del disegno (la forma della moneta è ellittica, la tavola è trapezoidale). Spesso è utile sovrapporre a ciò che si vuole copiare un'immaginaria griglia di linee verticali e orizzontali o una coppia di capelli incrociati, per aiutarsi a giudicare gli angoli reali delle linee osservate. Imparare a disegnare consiste in larga parte nell'imparare a scavalcare i normali processi della visione per rendere la propria esperienza dell'oggetto nel mondo *più simile al guardare un quadro*. Non potrà mai essere proprio come guardare un quadro, ma, una volta che la propria esperienza è stata alterata in quella direzione, si può, sfruttando altri trucchi del mestiere, «copiarla» più o meno bene sulla carta.

Si potrebbe pensare intuitivamente che il nostro campo visivo sia uniformemente dettagliato e messo a fuoco in ogni sua parte, dal centro fino alla periferia, ma un semplice esperimento mostra che non è così. Prendi un mazzo di carte da gioco ed estrai una carta coperta, in modo da non identificarla. Stendi il braccio e tieni la carta sulla periferia destra o sinistra del tuo campo visivo, poi girala, prestando attenzione a non volgere lo sguardo (scegli un punto di fronte a te e seguita a fissarlo). Scoprirai che non puoi dire neanche se è rossa o nera o se è coperta. Nota, tuttavia, che sei perfettamente consapevole di ogni pur minimo sobbalzo della carta. Stai osservando un

movimento senza essere in grado di vedere la forma o il colore della cosa che si muove. Ora inizia a muovere la carta verso il centro del tuo campo visivo, prestando di nuovo attenzione a non spostare lo sguardo. A quale punto puoi identificare il colore? A quale punto il seme e il numero? Nota che puoi dire se è una carta coperta o scoperta molto prima di poter dire se è un fante, una regina o un re. Sarai probabilmente sorpreso di quanto vicino al centro devi spostare la carta prima di essere in grado di identificarla.

Questa sbalorditiva deficienza della nostra visione periferica (in effetti tutto il campo visivo tranne due o tre gradi attorno al centro esatto) ci è normalmente nascosta dal fatto che i nostri occhi, a differenza delle cineprese, non sono saldamente puntati sul mondo ma sobbalzano qua e là in un gioco di inseguimento visivo, incessante e largamente non rilevato, con gli eventi potenzialmente interessanti che si verificano nel nostro campo visivo. O scivolando dolcemente o saltando con rapidi *movimenti saccadici*, i nostri occhi forniscono al cervello delle informazioni ad alta risoluzione su tutto ciò che occupa momentaneamente l'area foveale centrale del campo della retina. (La fovea dell'occhio ha una capacità discriminativa circa dieci volte superiore a quella delle aree circostanti della retina.)

Le nostre fenomenologie visive, i *contenuti* dell'esperienza visiva, sono in un formato dissimile da quello di qualunque altro modo di rappresentazione, non sono né quadri né film né enunciati né mappe né modelli in scala né diagrammi. Considera che cosa è presente nella tua esperienza quando osservi una folla sgomitante di migliaia di spettatori dall'altro lato di uno stadio. Non puoi identificare gli individui perché sono troppo lontani, a meno che qualche caratteristica macroscopica e ben marcata non ti aiuti (il presidente – sì, puoi dire che è proprio lui; capisci che è quello al centro del panneggio rosso, bianco e blu). Visivamente, puoi dire che la folla è composta di essere umani grazie al modo tipicamente umano con cui si muovono. C'è un qualcosa di globale nella tua esperienza visiva della folla (sembra tutto «affollato» lì, nello stesso modo in cui una macchia d'alberi visti attraverso la finestra può sembrare distintamente «ricca d'olmi» o un pavimento può sembrare «polveroso») ma tu non vedi una grande chiazza etichettata in qualche modo «folla»; vedi – contemporaneamente – migliaia di dettagli particolari: cappelli rossi altalenanti, occhiali scintillanti, pezzi di vestiti blu, programmi ondeggiati nell'aria, pugni alzati, ecc. Se tentassimo di rendere con una pittura «impressio-

nistica» la tua esperienza, la rissosa profusione di chiazze di colori *non* catturerebbe il contenuto; tu non hai l'esperienza di una rissosa profusione di chiazze di colori in questo caso, come non hai l'esperienza di un'ellissi quando osservi obliquamente una moneta. I dipinti – i quadri colorati in due dimensioni – possono avvicinarsi approssimativamente all'input retinico provocato da una scena tridimensionale e quindi possono creare un'impressione simile all'impressione visiva che si avrebbe vedendo la scena, ma allora il dipinto non sarebbe un dipinto *dell'*impressione risultante, ma piuttosto qualcosa che può provocare o stimolare una tale impressione.

Non si può dipingere un quadro realistico della fenomenologia visiva più di quanto non si possa farlo della giustizia o di una melodia o della felicità. Malgrado ciò sembra spesso appropriato, perfino irresistibile, parlare delle proprie esperienze visive come quadri nella testa. Questo fa parte della nostra fenomenologia visiva e quindi fa parte di ciò che deve essere spiegato nei successivi capitoli.

3. *La nostra esperienza del mondo interno*

Quali sono le «materie prime» della nostra vita interiore e come le utilizziamo? Non dovrebbe essere difficile trovare delle risposte: si può presumere che noi semplicemente «guardiamo e vediamo» e poi prendiamo nota dei risultati.

Secondo la tradizione ancora solida degli empiristi britannici, Locke, Berkeley e Hume, i sensi sono i varchi di accesso per l'arredamento mentale; una volta che sono sicuramente immagazzinati, questi materiali possono essere manipolati e combinati *ad libitum* per creare un mondo interiore di oggetti immaginari. Quando immagini una mucca viola che vola non fai altro che prendere il viola che hai ottenuto vedendo un ciclamino e le ali che hai ottenuto vedendo un'aquila, e le unisci alla mucca che hai ottenuto vedendo una mucca. Non è del tutto giusto. Ciò che penetra negli occhi è una radiazione elettromagnetica e non può quindi essere usato come una tinta con la quale dipingere mucche immaginarie. Varie forme di energia fisica bombardano i nostri sensi, subendo nei punti di contatto una «trasduzione» in impulsi nervosi che viaggiano verso il cervello. Quello che passa dall'esterno all'interno non è niente altro che informazione e benché la ricezione dell'informazione possa *provocare* la creazione di qualche oggetto fenome-

nologico (per esprimersi nel modo più neutrale possibile), è difficile credere che l'informazione stessa – che è soltanto un'astrazione concretizzata in qualche mezzo fisico modulato – possa *essere* l'oggetto fenomenologico. Ci sono ancora buone ragioni, tuttavia, per riconoscere, con gli empiristi britannici, che *in qualche maniera* il mondo interno debba dipendere dalle fonti sensoriali.

La vista è la modalità sensoriale più usata da noi pensatori umani per arricchire la nostra conoscenza percettiva, sebbene siamo pronti a ricorrere al tatto e all'udito per confermare ciò che i nostri occhi ci hanno detto. Questa nostra abitudine a rappresentarci il modo in cui la mente vede le cose attraverso la metafora della visione (un'abitudine che mi influenza per ben due volte nel periodo che stai leggendo) è una fonte considerevole di confusione e distorsione, come vedremo. La vista domina così ampiamente le nostre consuetudini intellettuali che abbiamo grandi difficoltà a concepire un'alternativa. Per raggiungere la comprensione di un problema, tracciamo dei diagrammi e degli schizzi ben visibili, così possiamo «vedere cosa succede» e se vogliamo «vedere se qualcosa è possibile» proviamo ad immaginarlo con il «nostro occhio mentale». Una razza di pensatori ciechi che si affidasse massicciamente all'udito sarebbe capace di comprendere con l'aiuto di motivetti, cantilene e guaiti nell'orecchio della mente tutto quello che noi riusciamo a capire grazie alle «immagini» mentali?

Anche coloro che sono affetti da una cecità congenita usano il vocabolario visivo per descrivere i propri processi mentali, benché non sia ancora chiaro in che misura ciò sia dovuto ad un ossequio nei confronti del linguaggio che imparano dalle persone vedenti o ad una riconosciuta appropriatezza della metafora malgrado le differenze dei loro processi mentali o perfino al fatto che usano il meccanismo visivo nel loro cervello in modo approssimativamente uguale a quello dei vedenti – malgrado siano privi delle normali porte d'ingresso. Rispondere a queste domande aiuterebbe a gettare luce sulla natura della coscienza umana normale, poiché il suo aspetto sostanzialmente visivo è uno dei suoi tratti caratteristici.

Noi abbiamo l'abitudine di segnalare la nostra comprensione di un problema dicendo che ne abbiamo «visto la soluzione» e questa non è meramente una vuota metafora. La natura apparentemente visiva della *fenomenologia* della comprensione è stata quasi completamente ignorata dagli studiosi di scienze cognitive, in particolar modo dai ricercatori nel campo

dell'Intelligenza Artificiale che hanno tentato di creare dei sistemi al calcolatore per la comprensione del linguaggio naturale. Perché hanno voltato le spalle alla fenomenologia? Molto probabilmente perché sono guidati dalla convinzione che la fenomenologia, per quanto reale e affascinante, non sia funzionale – equivale ad un ingranaggio che gira ma non svolge nessun compito nell'importante meccanismo della comprensione.

La fenomenologia di ascoltatori differenti di fronte alla stessa affermazione può variare quasi all'infinito senza alcuna evidente variazione nella loro comprensione. Considera le diverse immagini mentali evocate in due persone che ascoltano la stessa frase:

Ieri mio zio ha licenziato il suo avvocato.

Mario può incominciare a rappresentarsi vividamente la sua disavventura di *ieri*, inframmezzata da fugaci apparizioni di uno schema sulla relazione parentale (zio, cioè fratello della madre o del padre; o marito della sorella del padre o della madre), seguito poi da una scalinata di un palazzo di giustizia e un uomo anziano molto arrabbiato. Nello stesso tempo, Luisa è forse passata senza immagini sullo «ieri» e ha profuso la sua attenzione su qualche smorfia del volto di suo zio Giovanni, mentre le appare una porta sbattuta e l'uscita appena «intravista» di una donna elegantemente vestita etichettata «avvocato». Indipendentemente dalle loro immagini mentali sia Mario che Luisa hanno compreso ugualmente bene la frase precedente, come può essere confermato da una batteria di successive parafrasi e risposte a questionari. Inoltre, come qualche ricercatore più attento osserverà, l'immagine mentale *non può* essere la chiave della comprensione, perché non puoi fare un disegno di uno zio o dello ieri o del licenziare o di un avvocato. Gli zii, a differenza dei pagliacci e dei pompieri, non appaiono differenti in nessuna caratteristica che può essere visivamente rappresentata e gli ieri non assomigliano poi a un bel nulla. La comprensione, allora, non può essere realizzata da un processo che converte ogni cosa nella valuta delle immagini mentali, a meno che gli oggetti dipinti non siano identificati da specie di etichette, ma a quel punto la scrittura su tali etichette sarebbe una verbosità che necessita una comprensione, facendoci tornare al punto di partenza.

Io posso *udire* ciò che dici se lo dici ad una distanza suf-

ficientemente ravvicinata e se io sono cosciente. Io posso *comprendere* ciò che dici se si verificano varie circostanze, ma nessuna di queste sembra identificabile con qualche elemento della mia fenomenologia interiore; nessuna esperienza cosciente garantirà che io ti abbia capito o ti abbia frainteso. Il fatto che Luisa ha immaginato suo zio Giovanni non le impedisce minimamente di comprendere che è lo zio di colui che parla, non suo zio, che ha licenziato l'avvocato; lei *sa benissimo* ciò che l'interlocutore intendeva dire, stava solo raffigurandosi incidentalmente il volto dello zio Giovanni, con scarsissimo rischio di confusione, poiché la sua comprensione dell'interlocutore non dipende dalla sua immagine mentale.[4]

La comprensione, allora, non può essere spiegata chiamando in causa la connessa fenomenologia, ma ciò non significa che essa non sia realmente presente. In particolare non significa che un modello della comprensione che taccia sulla fenomenologia soddisfi le nostre ordinarie intuizioni sulla comprensione. Sicuramente una delle fonti più copiose che alimentano il diffuso scetticismo sulla «comprensione meccanica» dei linguaggi naturali consiste nel fatto che tali sistemi quasi mai si servono di un «laboratorio visivo» in cui osservano e analizzano le informazioni in ingresso. Se lo facessero, l'impressione che stiano realmente comprendendo ciò che elaborano acquisterebbe sicuramente forza (anche se rimarrebbe, come qualcuno insiste, solo un'illusione). Per il momento, se un calcolatore dicesse «Non vedo dove vuoi arrivare» in risposta a qualche input un po' ambiguo, ci sarebbe la forte tentazione di considerare una tale asserzione come un'ovvia frode.

La tentazione è certamente forte. Per esempio, è difficile immaginare come chiunque possa afferrare una barzelletta senza l'aiuto di qualche immagine mentale. Ci sono due amici al bar che stanno bevendo, uno si gira verso l'altro e dice: «Ehi, amico, penso che hai bevuto abbastanza: la tua faccia è tutta sfocata!». Ora, non hai usato un'immagine o un fuggevole schizzo per raffigurarti l'errore commesso da colui che parla? Questa esperienza ci fornisce un esempio, sembra, di *ciò che si sente quando si arriva a capire qualcosa*: ci sei tu, alle prese con un problema confuso, indecifrabile o almeno mai affrontato – qualcosa che in un modo o nell'altro ti provoca un prurito epistemico – quando finalmente *Aha, ci sono*! Sorge la luce, è arrivata la comprensione, l'oggetto è trasformato: diventa utile, capito, sotto controllo. Prima del tempo *t* la cosa non era capita; dopo il tempo *t*, è capita – si tratta di un cambiamento di stato

chiaramente definito che può essere spesso accuratamente cronometrato, benché sia una transizione scoperta introspettivamente e accessibile empaticamente e soggettivamente. È un errore, come vedremo, fare di questo esempio un modello per tutta la comprensione, ma è sicuramente vero che se il sorgere della comprensione ha una qualche fenomenologia, questa è la fenomenologia che ha.

Ci deve essere qualcosa di giusto nell'idea delle immagini mentali, e se è sbagliato pensarle come «fotografie nella testa», dovremo trovare qualche modo migliore per concepirle. L'immaginazione mentale non è limitata alla modalità visiva. Immagina la canzone *Bianco Natale*, stando attento a non canticchiarla come fai di solito; non riesci, tuttavia, a «udire» il motivo nel tuo orecchio mentale in una particolare tonalità? Se sei come me, ci riesci. Io non sono molto intonato, quindi non so dirti «dall'interno» in quale tonalità l'ho appena immaginata, ma, se qualcuno al piano suonasse *Bianco Natale*, immediatamente sarei in grado di dire con buona sicurezza o «Sì, questa è la tonalità con la quale l'ho immaginata» o «No, l'ho immaginata un'ottava più in alto».[5]

Non solo parliamo silenziosamente con noi stessi, ma talvolta lo facciamo con un particolare «tono di voce». Altre volte, sembra come se ci siano delle parole, ma non delle parole *udite*; ed altre volte ancora scopriamo che *ci sono* solo delle ombre o parvenze di parole a rivestire i nostri pensieri. Nei bei tempi della psicologia introspezionista, infuriava una disputa sulla possibilità che esista qualcosa come un pensiero *completamente* «privo di immagini». Per il momento possiamo lasciare aperta la questione, notando che alcune persone affermano recisamente di sì, e altre altrettanto recisamente di no. Nel prossimo capitolo appronteremo un metodo per dirimere controversie del genere. In ogni caso, la fenomenologia di un pensiero vivido non è limitata al *parlare* con se stessi; noi possiamo dipingere un quadro nel nostro *atelier* mentale, guidare una Ferrari, toccare una cravatta di seta o assaporare un immaginario tramezzino al salmone.

Queste sensazioni meramente immaginate (o ricordate) forse, come pensavano gli empiristi britannici, sono soltanto copie sbiadite delle nostre originarie sensazioni «che entrano dall'esterno», tuttavia esse possono provocare piacere o sofferenza esattamente come le sensazioni «reali». Come ogni sognatore ad occhi aperti sa, le fantasie erotiche non saranno un sostituto completamente soddisfacente della cosa reale, eppure

si sentirebbe sicuramente la loro mancanza se in qualche modo fossero fatte scomparire. Esse non solo recano piacere, ma possono anche far sorgere sensazioni reali ed altri ben conosciuti effetti corporei. Noi possiamo piangere mentre leggiamo un romanzo triste, e forse piangeva anche lo scrittore mentre lo scriveva.

Siamo tutti dei fini conoscitori delle gioie e dei dolori dell'immaginazione e molti di noi si considerano degli esperti nella preparazione di tali episodi che apprezziamo tanto, ma ci possiamo sorprendere tuttavia nell'apprendere quanto potente questa facoltà possa diventare con un serio allenamento. Io ho trovato sconvolgente, per esempio, che, quando si tengono dei concorsi di composizione musicale, i candidati spesso non sottomettono all'attenzione delle commissioni nastri o dischi (o esecuzioni dal vivo) delle loro opere, ma soltanto partiture scritte, e i giudici fanno tranquillamente le loro valutazioni *estetiche* sulla base della lettura delle partiture e dell'*ascolto* della *musica nella loro mente*. Quanto accurate possono essere le migliori immaginazioni musicali? Un musicista ben allenato può, leggendo solamente la partitura, dire in un battibaleno *esattamente* quale sarà il suono complessivo prodotto da quegli oboi e flauti in dissonanza rispetto agli strumenti ad arco? Possiamo citare una pletora di aneddoti, ma, per quanto ne sappia, questo è un campo ancora vergine in attesa di brillanti sperimentatori.

Le sensazioni immaginate (se ci è permesso chiamare così tali oggetti fenomenologici) costituiscono degli oggetti appropriati per l'apprezzamento e il giudizio estetico, ma perché, allora, le sensazioni reali ci stanno molto più a cuore? Perché non ci dovremmo accontentare di tramonti solamente ricordati, o di spaghetti al pesto solo pregustati? Gran parte delle gioie e dei dispiaceri associati agli eventi della nostra vita sono, dopo tutto, legati alla pregustazione o al rimpianto; le nude sensazioni rappresentano un fuggevole momento tra la gran massa delle cose che ci stanno a cuore. Perché – e come – le cose ci stanno a cuore costituirà l'argomento di prossimi capitoli, ma il fatto che le sensazioni immaginate, pregustate e ricordate siano piuttosto differenti dalle sensazioni *deboli* può essere facilmente chiarito con un altro auto-esperimento, che ci porta all'ingresso del terzo padiglione di fenom.

Adesso chiudi gli occhi e immagina che qualcuno con gli stivali rinforzati in ferro ti abbia assestato un violento calcio nello stinco sinistro, una decina di centimetri sopra la caviglia. Immagina il dolore lancinante che senti nel modo più dettagliato possibile; immagina che ti fa venire le lacrime agli occhi, immagina che sei quasi svenuto perché le fitte di dolore che senti sono così nauseanti che ti soverchiano. Lo hai appena immaginato molto vividamente; hai sentito dolore? Potresti a buon diritto lamentarti con me perché seguendo le mie istruzioni ti ho provocato un qualche dolore? Ho scoperto che la gente risponde in modo molto variegato a questo esercizio, ma nessuno mi ha ancora mai riferito che esso gli ha causato un dolore reale. Alcuni lo trovano inquietante, altri lo giudicano un esercizio mentale piuttosto piacevole, sicuramente meno sgradevole del più delicato pizzico nel braccio che chiameresti dolore.

Ora, supponi di aver sognato la stessa scena del calcio negli stinchi. Questo sogno potrebbe essere così scioccante da svegliarti; potresti perfino ritrovarti a massaggiare la gamba e a mugolare, con lacrime vere agli occhi. Ma non ci sarebbe nessuna infiammazione, nessun livido, nessuna contusione, e, non appena fossi sufficientemente sveglio e ti fossi abbastanza raccapezzato per formulare un giudizio sereno, diresti che il dolore non ha lasciato traccia sul tuo corpo – se mai c'è stato in qualche momento. I dolori sognati sono veri dolori o sono una sorta di dolori immaginari? O un qualcosa tra i due? E il dolore indotto dalla suggestione ipnotica?

Quantomeno, i dolori sognati, come quelli indotti dall'ipnosi, sono stati della mente che preferiamo evitare. Confrontali, comunque, con gli stati (della mente?) che nascono in te mentre stai dormendo e ti giri nel letto contorcendo inavvertitamente il braccio e poi, senza svegliarti, senza notarlo affatto, ti rigiri e assumi una posizione più confortevole. Sono dolori? Se fossi sveglio, gli stati causati in te da tali contorsioni sarebbero dolorosi. Ci sono delle persone, poche fortunatamente, che sono congenitamente insensibili al dolore. Prima che tu inizi ad invidiarli, dovresti sapere che, non correggendo la loro posizione durante il sonno (o anche da svegli!), diventano rapidamente storpi, poiché le loro articolazioni vengono danneggiate da un continuo abuso non limitato da campanelli d'allarme. Essi si bruciano, si tagliano e accorciano in mille modi la loro

vita infelice a causa di una manutenzione continuamente differita (Cohen ed altri, 1955; Kirman ed altri, 1968).

Non ci può essere alcun dubbio che avere un sistema d'allarme, composto di fibre del dolore e delle connesse porzioni del cervello, sia un grande dono evoluzionistico, anche se ciò significa che dobbiamo pagare il prezzo di *qualche* campanello che suona senza che ci possiamo fare nulla.[6] Ma perché i dolori devono *far tanto male*? Perché non potremmo avere semplicemente una sonora scampanellata nell'orecchio mentale, per esempio?

E qual è – se c'è – l'utilità della collera, della paura, dell'odio? (Presumo che l'utilità evoluzionistica della lussuria sia data per scontata.) Oppure, consideriamo un caso più complicato, la compassione. Rifletti sul significato etimologico di questa parola (e anche di «simpatia»): *patire insieme*. Puoi pensare alla *vibrazione simpatica*, quando la corda di uno strumento musicale è messa in vibrazione da un'altra vicina che è strettamente correlata alla prima perché entrambe condividono una stessa frequenza di risonanza naturale. Puoi immaginare una situazione in cui tuo figlio sia profondamente e crudelmente umiliato; non riesci a sopportarlo: sei inondato da emozioni travolgenti, ti senti affogare e con te la tua razionalità e la tua compostezza; stai per scoppiare a piangere, a litigare, a spaccare qualcosa. Questo sarebbe un caso estremo di compassione. Perché siamo fatti in modo da avere fenomeni del genere? Cosa sono?

Ci occuperemo della significanza adattativa (se tale si può valutare) dei vari stati affettivi in parecchi capitoli successivi; per il momento voglio solamente attirare l'attenzione, nella nostra passeggiatina introduttiva, sul consistente e innegabile contributo che gli affetti portano alla nostra convinzione che la coscienza sia importante. Considera il *divertimento*, per esempio. Tutti gli animali *vogliono continuare a vivere* – perlomeno combattono duramente per conservarsi nelle più diverse situazioni – ma solo poche specie riescono a colpirci per la loro capacità *di godersi la vita e divertirsi*. Ci possono venire in mente delle lontre vivaci che si rotolano nella neve, dei leoncini che si danno la caccia, i nostri cani e gatti domestici – non certo i pesci o i ragni. I cavalli, almeno i puledri, sembrano apprezzare il fatto di essere vivi, ma le mucche e le pecore di solito sembrano o annoiate o indifferenti. E non ti è mai capitato di pensare che il volare sia sprecato per gli uccelli, dal momento che non sembrano *apprezzare* tutta la delizia della loro attività? Il diverti-

mento non è un concetto banale, ma a mio avviso non ha ancora ricevuto l'attenzione che merita da parte dei filosofi. Certamente una spiegazione esauriente della coscienza dovrà illuminarci sul ruolo che essa ha nel permettere a noi (e solo a noi?) di divertirci. Ma come imparare a formulare le giuste domande? Un altro esempio ci aiuterà a capire le difficoltà che incontreremo.

Esiste una specie di primati nell'America Meridionale con abitudini più gregarie di altri mammiferi e con un curioso comportamento. I membri di questa specie spesso si riuniscono in gruppi più o meno grandi e nel corso dei loro chiacchiericci, per cause non identificate, vista la varietà di circostanze in cui il fenomeno si verifica, si trovano a subire degli attacchi di convulsioni respiratorie involontarie, una specie di ansito fragoroso e irresistibile, e che appare mutuamente rinforzato e così violento da mettere in crisi il loro normale funzionamento. Lungi dall'essere dannosi, comunque, tali attacchi sembrano desiderati dai membri di questa specie e alcuni di loro ne sembrano perfino dipendenti.

Potremmo essere tentati di pensare che se solo sapessimo che cosa si prova ad essere loro, se solo potessimo vederla dall'interno, noi capiremmo questa loro curiosa inclinazione. Se potessimo giudicarla «dal loro punto di vista», potremmo sapere a cosa serve. Ma in questo caso possiamo essere sicuri che una tale conoscenza non ci aiuterebbe a chiarire il mistero. Giacché noi abbiamo già l'accesso che cerchiamo: la specie è *homo sapiens* (che effettivamente abita il Sud America, ed altri luoghi) e il comportamento è il ridere.[7]

Nessun altro animale si comporta così. Un biologo che incontrasse un fenomeno talmente singolare dovrebbe dapprincipio chiedersi a che cosa potrebbe mai *servire*, ma poi, non riuscendo a formulare nessuna analisi plausibile dei diretti vantaggi biologici che potrebbe assicurare, sarebbe tentato di interpretare questo strano e non produttivo comportamento come un prezzo da pagare per qualche altro dono. Ma quale? Che cosa facciamo meglio grazie a questi meccanismi che implicano, come un prezzo che vale la pena di pagare, la nostra propensione – o quasi dipendenza – al riso? Il ridere riduce in qualche modo lo stress che accumuliamo durante le complicate attività intellettuali necessarie alla nostra vita socialmente avanzata? Perché, tuttavia, sono necessarie cose *divertenti* per ridurre lo stress? Perché non cose *verdi* o *cose semplicemente piatte*? Oppure, perché proprio *questo* comportamento è il prodotto

collaterale della riduzione dello stress? Perché non abbiamo una tendenza a bighellonare tremando e ruttando o a grattarci l'un l'altro la schiena o a canticchiare o a soffiarci il naso o a leccarci febbrilmente le mani?

Si noti che *il punto di vista interiore* è ben conosciuto e non problematico: ridiamo *perché ci divertiamo*, ridiamo perché le situazioni sono *divertenti* – e ridere è la risposta *appropriata* in una situazione divertente, mentre leccarsi le mani, per esempio, non lo sarebbe. È ovvio (infatti è *troppo* ovvio) perché ridiamo. Noi ridiamo perché siamo felici, perché siamo contenti, perché alcune cose sono comiche. Se volevamo un esempio di *virtus dormitiva* in una spiegazione, eccoci serviti: ridiamo a causa della comicità dello stimolo.[8] Questo è sicuramente vero; non ci sono altre ragioni perché noi ridiamo, quando ridiamo sinceramente. La comicità è la causa costitutiva del vero ridere. Proprio come il dolore è la causa costitutiva di un sincero comportamento addolorato. Poiché ciò è sicuramente vero, non dobbiamo negarlo.

Ma noi abbiamo bisogno di una spiegazione del ridere che vada oltre questa ovvia verità così come fanno le spiegazioni tipiche del dolore e del comportamento provocato dal dolore. Possiamo dare una descrizione biologica perfettamente valida del perché si deve verificare il dolore e il comportamento provocato dal dolore (infatti, l'abbiamo appena abbozzata); ciò che vogliamo è una descrizione altrettanto fondata del perché ci debbano essere la comicità e il ridere.

Possiamo sapere fin da ora, però, che, se ci presenteremo davvero con una simile descrizione, essa non soddisferà tutti! Alcune persone che si considerano *antiriduzioniste* lamentano che la descrizione biologica del dolore e del comportamento indotto dal dolore *lascia fuori la dolorosità*, «l'intrinseca spaventosità» del dolore che lo rende ciò che è; e presumibilmente esse avanzeranno le stesse riserve su qualsiasi descrizione del ridere che potremmo passare in rassegna: lascia fuori l'intrinseca comicità. Questa è la riserva tipica su tali spiegazioni: «Non hai fatto altro che spiegare il *comportamento conseguente* e il *meccanismo*, ma hai lasciato fuori la *cosa in sé*, cioè il dolore con tutta la sua spaventosità». Una tale posizione fa sorgere complicate questioni, che saranno discusse lungamente nel capitolo 12, ma per il momento limitiamoci a notare che qualsiasi descrizione del dolore che *lasciasse dentro* la spaventosità sarebbe circolare – presenterebbe una *virtus dormitiva* non scaricata. Similmente, un'appropriata descrizione del ridere deve lasciar fuori la pre-

sunta comicità intrinseca, l'allegria, la causticità, perché la loro presenza non farebbe altro che posticipare il tentativo di rispondere alla domanda.

La fenomenologia del ridere è sigillata ermeticamente: noi *capiamo* direttamente, naturalmente, senza deduzioni, con un'ovvietà che trascende l'intuizione, che ridere è ciò che accompagna la comicità – è la «giusta» reazione a uno stimolo umoristico. Possiamo provare a scomporre questa ovvietà atomica: la giusta reazione a una cosa buffa è il divertimento (uno stato interiore della mente); la naturale espressione del divertimento (quando non è importante nasconderlo o combatterlo, come talvolta accade) è il ridere. Sembra di aver ottenuto così quello che gli scienziati chiamerebbero una variabile intermedia, il divertimento, tra lo stimolo e la risposta, e sembra essere costitutivamente legata a entrambi gli estremi. Cioè, il divertimento è per definizione «ciò che provoca una sincera risata» ed è anche per definizione «ciò che è provocato da una cosa buffa». Tutto ciò è ovvio. Se le cose stanno in questi termini non sembra che abbiano bisogno di ulteriori spiegazioni. Come dice Wittgenstein: le spiegazioni devono a un certo punto arrestarsi. Ma tutto quello che abbiamo qui è un fatto bruto – ma precisamente spiegabile – della psicologia umana. Dobbiamo andare oltre la pura fenomenologia se vogliamo spiegare qualunque di questi abitatori del giardino fenomenologico.

Gli esempi di fenomenologia che abbiamo visto, malgrado le loro differenze, sembrano avere in comune due importanti caratteristiche. Da una parte, costituiscono le nostre conoscenze più intime: nulla può essere conosciuto meglio degli oggetti della nostra personale fenomenologia – o perlomeno così sembra. D'altra parte, sono provocatoriamente inaccessibili alla scienza materialistica: nulla può essere meno simile a un elettrone o una molecola o un neurone *del modo in cui il tramonto appare a me in questo momento* – o perlomeno così sembra. I filosofi sono stati giustamente impressionati da entrambe le caratteristiche e hanno enfatizzato in vari modi l'aspetto che a loro sembrava più problematico. Per alcuni, il grande enigma è costituito dalla speciale intimità della fenomenologia: come è possibile che il nostro giudizio sia *insindacabile*, che si basi su un *accesso privilegiato* o una *conoscenza diretta* di tali oggetti? Qual è la differenza tra le nostre relazioni epistemiche con la nostra fenomenologia e le nostre relazioni epistemiche con gli oggetti del mondo esterno? Per altri, il grande enigma riguarda le inusitate «qualità intrinseche» – o, per usare la parola latina, i

qualia – della nostra fenomenologia: come è possibile che qualcosa composto di particelle materiali *sia* la gioia che provo, o sia «in ultima analisi omogeneo» (Sellars, 1963) al cubetto di ghiaccio rosa che sto adesso immaginando, o mi riguardi nello stesso modo in cui mi riguarda il mio dolore?

Trovare una descrizione materialistica che renda giustizia a tutti questi fenomeni non sarà facile. Abbiamo fatto progressi, però. Il nostro breve inventario ha incluso qualche caso in cui un po' di conoscenza dei meccanismi soggiacenti sfida – e forse perfino usurpa – l'autorità che usualmente garantiamo a ciò che è ovvio introspettivamente. Avvicinandosi un po' di più agli oggetti in esposizione, e osservandoli da prospettive differenti, abbiamo iniziato a rompere l'incantesimo, a dissipare la «magia» che avvolge il giardino fenomenologico.

4

UN METODO PER LA FENOMENOLOGIA

1. *La prima persona plurale*

Non si fa della seria zoologia semplicemente passeggiando nello zoo, guardando questo e quello, e meravigliandosi delle curiosità. Una zoologia seria richiede precisione, che si ottiene disponendo di metodi unanimemente accettati di descrizione ed analisi, in modo che i vari zoologi possano essere sicuri di capirsi quando si parlano. Una fenomenologia seria ha ancora più bisogno di un metodo chiaro e neutrale di descrizione, perché sembra proprio che ognuno usi le parole dotandole di un significato personale e per di più sentendosi un esperto. È davvero sorprendente notare la frequenza con la quale le discussioni «accademiche» sulle controversie fenomenologiche degenerino in assurde cacofonie prodotte da pugni sulla scrivania e da parole disperse nel vento. Ciò è in un certo senso ancor più stupefacente poiché secondo una consolidata tradizione filosofica, *noi concordiamo tutti* su ciò che scopriamo quando «guardiamo dentro» la nostra fenomenologia.

L'indagine fenomenologica è stata di solito considerata una pratica collettiva molto affidabile, un mettere in comune delle osservazioni condivisibili. Quando Cartesio scrisse le sue *Meditazioni* come un soliloquio in prima persona singolare, egli si aspettava chiaramente che i suoi lettori arrivassero a condividere ognuna delle sue osservazioni, riproducendo nella loro mente le indagini da lui proposte e ottenendo gli stessi risultati. Anche gli empiristi britannici, Locke, Berkeley e Hume, scrivevano presumendo che ciò che stavano facendo, la maggior parte del tempo, era dell'*introspezione*, e che la loro introspezione sarebbe stata facilmente replicata dai loro lettori. Locke espresse chiaramente questo presupposto nel suo *Saggio sull'intelletto umano* (1690) chiamando il suo metodo «un semplice metodo storico» – privo di astruse deduzioni o teorizzazioni *a*

priori, una semplice descrizione dei fatti osservati, per mostrare ai lettori ciò che sarebbe stato manifesto a tutti con un po' di osservazione. In effetti, quasi tutti gli autori che hanno scritto sulla coscienza si sono basati su ciò che potremmo chiamare il *presupposto della prima persona plurale*: quali che siano i misteri della coscienza, *noi* (tu, gentile lettore, ed io) possiamo tranquillamente parlare delle nostre comuni conoscenze, delle cose che entrambi troviamo nei nostri flussi di coscienza. E, con poche riluttanti eccezioni, i lettori hanno sempre accettato di far parte della cospirazione.

Questo andrebbe bene, se non ci fosse il fatto imbarazzante che le affermazioni fatte in queste condizioni di educato e reciproco accordo sono infestate da controversie e contraddizioni. Noi ci stiamo ingannando su qualcosa. Forse ci stiamo ingannando sul grado in cui noi siamo tutti fondamentalmente uguali. Forse, quando una persona per la prima volta entra in contatto con le differenti scuole di pensiero sulla fenomenologia, sceglie la scuola che a lei sembra più giusta; e ogni scuola di descrizione fenomenologica è fondamentalmente nel giusto sul tipo di vita interiore dei propri membri, ma poi inavvertitamente arriva a delle generalizzazioni che pretendono erroneamente di dire come stanno le cose per tutti.

O forse ci stiamo ingannando sull'alta affidabilità dell'introspezione, dei nostri poteri di auto-osservazione della nostra mente cosciente. Fin dai tempi di Cartesio e del suo «*cogito ergo sum*», questa nostra capacità è stata sempre considerata in un certo senso immune da errore; noi abbiamo un accesso privilegiato ai nostri pensieri e ai nostri sentimenti, un accesso che è sicuramente migliore di quello di un estraneo. («Immagina qualcuno che provi a dirti che ti stai sbagliando su ciò che stai pensando o sentendo!») Noi siamo o «infallibili» – sempre sicuramente nel giusto – o perlomeno «incorreggibili» – giusto o sbagliato, nessun altro può correggerci (Rorty, 1970).

Ma forse questa dottrina dell'infallibilità, per quanto consolidata, è solo un errore. Forse anche se *siamo* tutti fondamentalmente simili nella nostra fenomenologia, alcuni osservatori fanno semplicemente confusione quando provano a descriverla, ma, poiché sono così sicuri di essere nel giusto, sono relativamente impermeabili alle correzioni (sono incorreggibili nel senso dispregiativo del termine). In ogni modo, le controversie sono reali. E c'è ancora un'altra possibilità che credo sia più vicina alla verità: ciò su cui ci inganniamo è l'idea che l'attività di «introspezione» sia *sempre* una questione di semplice «osser-

vazione». Sospetto che quando affermiamo di star usando i nostri poteri solo per *guardarci dentro*, in realtà siamo sempre impegnati in una specie di *teorizzazione* improvvisata – e siamo dei teorici molto ingenui perché non ci rendiamo conto che c'è così *poco* da «osservare» e così tanto su cui pontificare senza paura di essere contraddetti. Quando facciamo dell'introspezione, comunemente, ci troviamo davvero molto vicini alla situazione di quel leggendario cieco che sta esaminando le diverse parti dell'elefante. Questa sembra a prima vista un'idea strampalata, ma vediamo cosa può essere detto in suo favore.

Qualcosa che abbiamo incontrato nel nostro giro attraverso fenom ti ha colpito? Non sei stato sorpreso, per esempio, dal fatto che non puoi identificare una carta da gioco fino a che non è quasi perfettamente al centro del tuo campo visivo? Trovo che la maggior parte delle persone rimangono sorprese – anche quelle che sanno che la capacità di risoluzione della nostra visione periferica è limitata. Se ti ha sorpreso, ciò deve significare che se tu avessi fatto delle affermazioni sull'argomento prima della dimostrazione sorprendente, avresti molto probabilmente detto qualcosa d'impreciso. Spesso noi pretendiamo di avere una conoscenza diretta di una *quantità di contenuti* nel nostro campo visivo più ampia di quella che in realtà abbiamo. Perché questa pretesa ci sembra così convincente? Non perché abbiamo osservato direttamente e «incorreggibilmente» noi stessi mentre esperiamo questo contenuto periferico, ma perché ciò *sembra ragionevole*. Dopo tutto, non notiamo nessun intervallo nero nel nostro campo visivo in condizioni normali e sicuramente, se ci fosse un'area che non fosse ben colorata, noteremmo la discrepanza, e inoltre dovunque ci rivolgiamo vediamo ogni cosa colorata e ben dettagliata. Se pensiamo che il nostro campo visivo soggettivo sia fondamentalmente un quadro interno composto da forme colorate, allora è ragionevole pensare che ogni porzione della tela abbia *qualche* colore – anche la tela grezza ha *un* colore! Ma questa è una conclusione tratta da un modello opinabile del nostro campo visivo soggettivo, non da qualcosa che abbiamo osservato.

Sto forse dicendo che non abbiamo alcun accesso privilegiato alla nostra esperienza cosciente? No, sto dicendo solamente che abbiamo la tendenza a ritenerci più immuni da errori di quanto non siamo. Molte persone normalmente ammettono, se messe alle strette in questo modo sul loro accesso privilegiato, che non hanno alcun accesso speciale *alle cause e agli effetti* della loro esperienza cosciente. Per esempio, possono essere sorprese

nell'apprendere che gustano con il naso o ascoltano le note molto basse con i piedi, ma non avevano mai preteso di essere un'autorità sulle fonti o le cause della loro esperienza. Essi sono l'autorità, dicono, solo sulle esperienze stesse, separate dalle cause e dagli effetti. Ma anche se molta gente *dice* di pretendere l'autorità solo sui contenuti isolati delle proprie esperienze, e non sulle loro cause ed effetti, spesso supera questi limiti autoimposti. Per esempio, saresti pronto a scommettere sulle seguenti affermazioni? (Almeno una di esse è stata inventata da me.)

1. Puoi vedere una macchia di colore che è completamente e contemporaneamente sia rossa che verde – una macchia di *entrambi* i colori (non mischiati) nello stesso momento.
2. Se osservi un cerchio giallo su uno sfondo blu (con una buona luce) e la luminanza o la brillantezza del giallo e del blu sono ben equilibrate, il confine tra il giallo e il blu scompare.
3. Esiste un suono, chiamato talvolta «spirale acustica senza fine», la cui altezza sembra crescere sempre più, ma che in realtà non diventa affatto più alto.
4. Esiste un'erba che in dosi eccessive non ti fa capire un discorso nella tua lingua nativa. Fino a che l'effetto non sia scomparso, riesci ad ascoltare tutto perfettamente, senza brusii o disturbi, ma le parole che ascolti ti suonano come una lingua straniera, anche se in un certo senso sai che non è così.
5. Se quando sei bendato viene applicato un vibratore in un punto particolare del tuo braccio mentre ti tocchi il naso, senti che il tuo naso cresce come quello di Pinocchio; e se il vibratore è spostato su un altro punto, avrai allora la strana sensazione di spingere il tuo naso verso l'interno, con l'indice che arriva fin dentro il tuo cranio.

Ho inventato l'affermazione numero 4, sebbene, per quanto mi risulta, potrebbe essere vera. Dopo tutto, a causa di una ben studiata neuropatologia chiamata prosopagnosia, non è possibile riconoscere le facce degli amici e conoscenti più vicini, ma la vista è sostanzialmente intatta ed è facile identificare la maggior parte delle cose.[1] Il punto che voglio sottolineare è, di nuovo, non che *non* hai un accesso privilegiato alla natura o al contenuto delle tue esperienze coscienti, ma solo

che dovremmo stare attenti alla tendenza *molto* allettante a sopravvalutare tale accesso.

Durante la nostra visita guidata a fenom, ho proposto di effettuare una serie di semplici esperimenti; ma questo non era in linea con lo spirito della fenomenologia «pura». I fenomenologi, infatti, tendono a sostenere che, *poiché* non abbiamo autorità sulle cause e gli effetti fisiologici della nostra fenomenologia, noi dovremmo ignorare tali cause ed effetti nel nostro tentativo di dare una descrizione pura, neutrale e preteoretica di ciò che ci viene «dato» nel corso della nostra esperienza quotidiana. Forse hanno ragione, ma allora pensa a quanti affascinanti abitatori di fenom non avresti mai incontrato! Uno zoologo che tentasse di estrapolare l'intera scienza dall'osservazione di un cane, un gatto, un cavallo e un pesciolino rosso perderebbe probabilmente qualcosa.

2. *La prospettiva in terza persona*

Poiché indulgeremo nella fenomenologia *impura*, dobbiamo essere estremamente precisi sul metodo. La prospettiva usualmente adottata dai fenomenologi è quella di Cartesio, *la prospettiva in prima persona*, nella quale *io* descrivo in un monologo (che *ti* lascio ascoltare) quello che *io* trovo nella *mia* esperienza cosciente, contando sul fatto che tutti *noi* saremo d'accordo. Ho cercato di mostrare, comunque, che l'intima complicità stimolata dalla *prospettiva in prima persona plurale* è un'infida incubatrice di errori. Nella storia della psicologia, infatti, proprio il crescente riconoscimento di questo problema metodologico ha condotto al declino dell'introspezionismo e al sorgere del comportamentismo. I comportamentisti decisero di evitare meticolosamente ogni congettura su ciò che accade nella *mia* o *tua* o *sua* mente e difesero *la prospettiva in terza persona*, nella quale solo i fatti ricavati «dall'esterno» contano come dati. Puoi filmare la gente in azione e poi misurare le percentuali di errori commessi in compiti che implicano movimenti corporei o i tempi di reazione necessari a premere pulsanti o abbassare leve oppure puoi misurare il ritmo cardiaco o le onde cerebrali o i movimenti degli occhi o (se hai una macchina che li quantifichi oggettivamente) i rossori della presunta vergogna o i cambiamenti di potenziale elettrico della pelle (quelli rilevati dalle «macchine della verità»). Puoi aprire il cranio delle persone (chirurgicamente o con congegni per l'analisi cerebrale) per

vedere cosa accade nel loro *cervello*, ma non devi fare nessuna *supposizione* su ciò che accade nella loro *mente*, perché su di essa non potrai mai ottenere nessun dato fintanto che usi i metodi intersoggettivamente verificabili della scienza fisica.

Il nocciolo dell'idea è questo: poiché non puoi mai «vedere direttamente» nella mente della gente, ma ti devi fidare delle loro parole, tutti i possibili fatti sugli eventi mentali non possono rientrare nei dati della scienza, poiché non possono mai essere verificati adeguatamente con metodi oggettivi. Questo scrupolo *metodologico*, che guida oggi *tutta* la neuroscienza e la psicologia sperimentale (non solo la ricerca «comportamentista»), è stato troppo spesso innalzato ad un principio *ideologico*, come uno dei seguenti:

- Gli eventi mentali non esistono. (Punto e basta! – questo è stato giustamente chiamato «comportamentismo selvaggio».)

- Gli eventi mentali esistono ma non producono alcun effetto, quindi la scienza non può studiarli (epifenomenismo – vedi capitolo 12, paragrafo 5).

- Gli eventi mentali esistono e producono effetti ma *questi* effetti non possono essere studiati dalla scienza, che quindi si deve accontentare di teorie relative agli effetti e ai processi «periferici» o «inferiori» del cervello. (Questa opinione è molto diffusa tra i neuroscienziati, specialmente quelli che guardano con sospetto i «teorici». In realtà è una forma di dualismo; questi ricercatori concordano evidentemente con Cartesio che la mente *non* è il cervello e sono disposti ad avere una teoria che riguarda solo il cervello.)

Tutte queste concezioni saltano ingiustificatamente alle conclusioni. Anche se gli eventi mentali non sono tra i *dati* della scienza, non significa che non siano studiabili scientificamente. I buchi neri e i geni non sono tra i dati della scienza, ma noi sappiamo sviluppare delle buone teorie scientifiche che li riguardano. La sfida consiste nell'elaborare una teoria degli eventi mentali usando i dati che il metodo scientifico permette.

L'elaborazione di una simile teoria deve basarsi, come *tutta* la scienza, su un punto di vista in terza persona; ma

alcune persone diranno che una tale teoria della mente cosciente non è possibile. Tra i più notevoli, il filosofo Thomas Nagel ha sostenuto che

> Ci sono cose sul mondo e la vita e noi stessi che non possono essere adeguatamente comprese da un punto di vista completamente oggettivo, per quanto esso possa far progredire la nostra comprensione oltre il punto dal quale siamo partiti. Troppe cose sono essenzialmente connesse ad un particolare punto di vista o tipo di punto di vista, e quindi ogni tentativo di dare una spiegazione completa del mondo in termini oggettivi e distaccati da quelle prospettive conduce inevitabilmente a false riduzioni o franche negazioni dell'esistenza di certi fenomeni che sono invece chiaramente reali. (Nagel, 1986, p. 7)

Vedremo. È prematuro discutere su ciò che può essere o non può essere spiegato da una teoria fino a che non si vede che cosa dice davvero la teoria. Ma se dobbiamo ascoltare senza pregiudizi ciò che una teoria può dirci abbiamo bisogno, dato che ci troviamo di fronte ad un tale scetticismo, di un modo neutrale di *descrivere i dati* – un metodo che non pregiudica la questione. Sembrerebbe che un metodo neutrale di questa specie non possa esistere, ma non è vero e io prima lo descriverò e poi lo adotterò.

3. *Il metodo dell'eterofenomenologia*[2]

Il termine è sinistro; non semplice fenomenologia ma *etero*fenomenologia. Che sarà mai? In realtà, è qualcosa di familiare a tutti noi, sia scienziati che profani, ma che va presentato con estrema cautela, notando ciò che presuppone e ciò che implica, poiché assumendo questo metodo si compie un passo teoretico gigantesco. Ignorando tutte le invitanti scorciatoie, possiamo allora servirci di un sentiero *neutrale* che ci conduce dalla scienza fisica oggettiva, e dalla sua insistenza sulla prospettiva in terza persona, ad un metodo per la descrizione fenomenologica che può (in linea di principio) rendere giustizia delle esperienze soggettive più private e ineffabili pur senza mai abbandonare gli scrupoli metodologici della scienza.

Vogliamo avere una teoria della coscienza, ma non tutti concordano su quali entità abbiano una coscienza. La hanno gli esseri umani appena nati? E le rane? E le ostriche, le formiche, i robot, gli zombi ...? Dovremmo restare neutrali per il momento

su questo, ma esiste una classe di entità a cui quasi tutti attribuiscono una coscienza, e cioè i nostri simili, gli esseri umani adulti.

Ora, alcuni di questi esseri umani adulti *potrebbero* essere degli zombi – nel senso «tecnico» dei filosofi. Il termine *zombi* viene dalla cultura vudu haitiana e si riferisce, in quel contesto, a un «morto vivente», una persona punita per qualche misfatto e condannata a vagare, biascicando e con occhi cadaverici, obbedendo meccanicamente agli ordini di qualche sciamano o prete vudu. A tutti è capitato di vedere degli zombi nei film dell'orrore, ed è chiaro che sono immediatamente distinguibili dalla gente normale. (Gli zombi haitiani, grosso modo, non possono danzare, raccontare barzellette, sostenere delle animate discussioni filosofiche, partecipare a una brillante conversazione – e inoltre sono orrendi.)[3] Ma i filosofi usano il termine *zombi* per riferirsi ad una categoria differente di esseri umani immaginari. Per un comune accordo tra i filosofi, uno zombi è o sarebbe un essere umano che si comporta in modo perfettamente normale, è sveglio, loquace e vivace come tutti i suoi simili, ma in realtà non è affatto cosciente, è piuttosto una specie di automa. Ciò che è importante nella nozione filosofica di zombi è che non si può distinguere uno zombi da una persona normale osservando solamente il suo comportamento esteriore. Poiché questo è tutto quello che possiamo vedere dei nostri amici e vicini, *alcuni dei tuoi migliori amici potrebbero essere zombi.* Questa è, perlomeno, la tradizione, e quindi è bene che per il momento io mi mantenga neutrale. Così, sebbene il metodo che propongo non faccia alcun assunto sulla *reale coscienza* di nessun essere umano adulto apparentemente normale, si focalizza proprio su questa classe di esseri umani adulti e normali, poiché, se esiste davvero la coscienza, la troveremo senz'altro tra loro. Dopo aver delineato, almeno in generale, una teoria della coscienza umana potremo rivolgere la nostra attenzione alla (eventuale) coscienza delle altre specie, inclusi gli scimpanzè, i delfini, le piante, gli zombi, i marziani e i tostapane (i filosofi hanno spesso l'abitudine di indulgere in fantasticherie durante i loro esperimenti mentali).

Gli esseri umani adulti vengono studiati da molte scienze. I loro corpi vengono sondati dai biologi e dai medici, dai dietologi e dagli ingegneri (che si chiedono, per esempio, «Qual è la velocità massima che può raggiungere il dito di un dattilografo?» o «Qual è la resistenza alla trazione di un capello umano?»). Essi vengono studiati anche dagli psicologi e dai

neuroscienziati, che pongono le singole persone, in questo caso i *soggetti*, in varie situazioni sperimentali. Nella maggior parte degli esperimenti, i soggetti devono essere catalogati e preparati; si devono annotare tutte le loro particolarità interessanti, per esempio l'età, il genere, il grado d'istruzione, se mancini o no, ecc., ma soprattutto *si deve dire loro cosa fare*. Questa è la differenza più importante tra i soggetti umani e, diciamo, le colture virali dei biologi o i campioni di materiali speciali degli ingegneri o le soluzioni dei chimici o i topi, gatti e piccioni degli psicologi animali.

Le persone sono gli unici oggetti di uno studio scientifico la cui preparazione implica generalmente (ma non sempre) la comunicazione verbale. Ciò è dovuto in parte ad un problema di etica della scienza: la gente non può essere usata negli esperimenti senza il suo consenso e non è possibile ottenere un consenso esplicito senza l'interazione verbale. Ma è dovuto anche, e ciò è più importante dal nostro punto di vista, al fatto che attraverso la comunicazione verbale si possono stabilire le condizioni e le regole degli esperimenti. Ai soggetti viene chiesto di eseguire vari compiti intellettuali, risolvere problemi, osservare oggetti su schermi, premere pulsanti, fare valutazioni, eccetera. La validità della maggior parte degli esperimenti dipende da una preparazione adeguata e uniforme; se, per esempio, le istruzioni a un soggetto che parla solo inglese vengono date in turco, allora il fallimento dell'esperimento è assicurato; ma anche fraintendimenti meno rilevanti possono compromettere gli esperimenti, ed è quindi essenziale che questa preparazione dei soggetti umani tramite la comunicazione verbale sia fatta con la massima accuratezza.

Che cosa implica questa prassi di parlare con i soggetti? È sicuramente un elemento non eliminabile degli esperimenti psicologici, ma non presuppone forse la coscienza dei soggetti? Gli sperimentatori non finiscono così in compagnia degli introspezionisti, dovendo fidarsi della parola incontrollabile del soggetto su ciò che egli o ella capisce? Non corriamo il rischio di essere ingannati da zombi o robot o altri impostori?

Dobbiamo osservare più da vicino i dettagli di un generico esperimento con soggetti umani. Supponiamo, come spesso accade, che l'intero esperimento sia «catturato» da una registrazione multipla: una videoregistrazione, una registrazione del suono, dell'elettroencefalogramma, e così via. Nulla che non sia così registrato conterà come dato. Concentriamoci sulla registrazione dei suoni – soprattutto vocali – prodotti dai

soggetti e dagli sperimentatori nel corso dell'esperimento. Poiché i suoni prodotti dai soggetti sono prodotti con mezzi fisici, essi sono in linea di principio spiegabili e prevedibili dalla fisica, usando gli stessi princìpi, le stesse leggi e gli stessi modelli che usiamo per spiegare e prevedere i rumori o gli scoppi di un motore d'automobile. Oppure, poiché i suoni sono prodotti da mezzi fisiologici, possiamo aggiungere i principi della fisiologia e tentare di spiegare i suoni usando le risorse di quella scienza, proprio come facciamo con il russare, i singulti, i gorgoglii di stomaco e gli scricchiolii delle articolazioni. Ma i suoni cui siamo principalmente interessati sono, ovviamente, quelli vocali e più in particolare il loro sottoinsieme formato dai suoni vocali che sono *apparentemente* suscettibili di analisi linguistica o semantica (escludendo quindi starnuti, rutti e sbadigli fortuiti). Non è sempre ovvio stabilire esattamente quali suoni includere in questo sottoinsieme, ma c'è un modo per allontanare gli errori: diamo delle copie della registrazione a tre esperti stenografi e chiediamo a ognuno di fornirci la *sua trascrizione* dei dati grezzi.

Questo semplice passo è carico di implicazioni; con esso ci spostiamo da un mondo – quello dei meri suoni fisici – a un altro: il mondo delle parole e dei significati, della sintassi e della semantica. Questo passo produce una ricostruzione radicale dei dati, un'astrazione dalle loro proprietà acustiche o genericamente fisiche a una successione di parole (sebbene ancora dotata di una precisa temporalizzazione – vedi, per esempio, Ericsson e Simon, 1984). Quali criteri guidano tale ricostruzione? Sebbene si possa presumere che ci siano delle relazioni regolari e rilevabili tra le proprietà fisiche delle onde sonore registrate sul nastro e i *fonemi* che un dattilografo sente e poi trascrive in parole, noi non siamo ancora in grado di descriverle in dettaglio. (Se così fosse, il problema di una macchina in grado di scrivere sotto dettatura sarebbe risolto; molti progressi sono stati fatti in questo campo, ma ci sono ancora delle grandi difficoltà.) In attesa dei risultati di queste ricerche acustiche e fonologiche, possiamo sempre fidarci delle nostre trascrizioni e considerarle delle traduzioni oggettive dei dati fintanto che prendiamo alcune precauzioni elementari. In primo luogo, affidare la trascrizione a degli stenografi (invece di far fare questo lavoro allo sperimentatore stesso) ci preserva da pregiudizi o forzature interpretative sia volontarie che involontarie. (Gli stenografi nelle aule di giustizia svolgono lo stesso ruolo neutrale.) Inoltre, le tre trascrizioni indipendenti ci danno una

buona garanzia dell'oggettività del processo. Presumibilmente, se la registrazione è buona, le trascrizioni coincideranno parola per parola tranne una lievissima porzione inferiore all'un per cento del testo. Per questi rari casi di disaccordo potremmo o semplicemente scartare i dati o fidarci della concordanza di almeno due trascrizioni per stabilire il risultato definitivo.

La trascrizione o il testo non è, strettamente parlando, un *dato*, giacché, come abbiamo visto, è ottenuto sottoponendo i dati grezzi a un processo di interpretazione. Questo processo interpretativo dipende da presupposizioni che riguardano la lingua impiegata e le intenzioni del parlante. Per chiarire questo punto basta paragonare il compito che abbiamo affidato agli stenografi con il compito di trascrivere una registrazione di un canto di uccelli o di un grugnito di maiali. Quando un essere umano parla, lo stenografo (ma anche un bravo scolaro delle elementari) è in grado di trascrivere i suoni uditi in frasi scritte; ma solo se conosce la lingua del parlante; e se, per esempio, il soggetto dice «Adesso l'oggetto si muove da dinistra a sestra» noi permetteremo agli stenografi di correggerlo in «Adesso l'oggetto si muove da sinistra a destra». Nessuna strategia di purificazione similare è disponibile per le trascrizioni dei canti degli uccelli o dei grugniti dei maiali – almeno fino a quando qualche ricercatore non scoprirà che esistono delle regole nella produzione di tali rumori ed escogiterà un sistema descrittivo e normativo.

Senza alcuno sforzo – in effetti, involontariamente – noi «diamo un senso» al flusso sonoro trasformandolo in parole (a questo proposito, faremo bene a permettere agli stenografi di cambiare «da dinistra a sestra» in «da sinistra a destra», perché probabilmente lo cambierebbero senza accorgersene). Il fatto che il processo sia altamente affidabile e quasi inosservabile in normali circostanze non dovrebbe nasconderci il fatto che è un processo sofisticato anche quando non arriva fino alla comprensione ma si arresta al riconoscimento delle parole. Quando lo stenografo trascrive «Mi sembrava che ci fosse una risonante sensazione di partecipazione nel mio presentimento, un significante flusso implicito di attrazioni e ripulse, un complesso sovrapporsi di conferme anticipatrici che rivelavano superfici dietro superfici», può non avere la minima idea di quello che ciò significa, ma essere piuttosto certo che queste erano effettivamente le parole che l'oratore intendeva pronunciare, ed è riuscito a pronunciare, indipendentemente da ciò che significano.

È sempre possibile che anche l'oratore non conosca il significato delle parole. Il soggetto, dopo tutto, *potrebbe* essere uno zombi o un pappagallo travestito da persona o un calcolatore che dirige un sintetizzatore. Oppure, più plausibilmente, il soggetto può essersi confuso o essere in preda a qualche teoria mal compresa o semplicemente sta prendendo in giro lo sperimentatore sparando una serie di assurdità. Per il momento voglio sottolineare che il processo che conduce alla creazione di una trascrizione o un testo dai dati registrati è neutrale rispetto a tutte queste strane possibilità, anche se presuppone metodologicamente che c'è un testo da ricavare. Quando non si può ricavare nessun testo, faremmo bene a gettar via la registrazione relativa a quel soggetto e a ricominciare da capo.

Fin qui il metodo descritto non è particolarmente originale o controverso: abbiamo raggiunto la banale conclusione che possiamo trasformare una registrazione su un nastro in un testo senza abbandonare la scienza. Abbiamo proceduto con estrema cautela nello stabilire questo risultato poiché il prossimo passo, da una parte, permette di studiare la coscienza empiricamente, dall'altra, crea molti ostacoli e confusione. Dobbiamo andare oltre il testo; dobbiamo interpretarlo come un documento attestante degli *atti linguistici*; non semplici emissioni vocali o recitazioni ma asserzioni, domande, risposte, promesse, commenti, richieste di chiarificazione, meditazioni ad alta voce, autoammonizioni.

Questo tipo di interpretazione ci richiede di adottare quello che io chiamo l'*atteggiamento intenzionale* (Dennett, 1971, 1978a, 1987a): dobbiamo trattare colui che emette i rumori come un agente, un agente razionale, in cui albergano credenze e desideri e altri stati mentali che mostrano *intenzionalità* o «riferimento», e le cui azioni possono essere spiegate (o previste) basandosi sul contenuto di quegli stati. Allora i suoni prodotti devono essere interpretati, per esempio, come cose che i soggetti *volevano dire*, *proposizioni* che intendevano *asserire*, per varie *ragioni*. In realtà, stavamo già basandoci su un'assunzione del genere quando abbiamo purificato il testo (abbiamo ragionato così: perché mai qualcuno *vorrebbe dire* «da dinistra a sestra»?).

Quali che siano i pericoli che corriamo nell'adottare l'atteggiamento intenzionale nei confronti di questi comportamenti verbali, essi costituiscono il prezzo che dobbiamo pagare per accedere a una serie di affidabili ovvietà che sfruttiamo nella strutturazione degli esperimenti. Ci sono molti motivi per voler dire qualcosa ed è importante escluderne alcuni tramite la

progettazione sperimentale. Talvolta, per esempio, la gente vuole dire qualcosa non perché ci crede ma perché crede che il suo ascoltatore vuole sentirsela dire. È di solito importante intraprendere i passi necessari per diminuire la probabilità che questo desiderio sia presente ed effettivo: diciamo ai soggetti che quello che vogliamo sentire è *qualsiasi cosa essi credano* e ci preoccupiamo di non lasciar trapelare ciò che desideriamo essi credano. In altre parole, facciamo del nostro meglio per metterli in una situazione in cui, dati i desideri che abbiamo istigato in loro (il desiderio di cooperare, di essere pagato, di essere un buon soggetto), essi non abbiano alternative migliori di quella di dire ciò che effettivamente credono.

L'atteggiamento intenzionale nei confronti dei nostri soggetti è necessario anche se vogliamo servirci di quell'utile classe di eventi costituita dalla pressione di un pulsante da parte dei soggetti. Di solito, premere un pulsante equivale ad eseguire un qualche atto linguistico convenzionalmente stabilito, come *asserire* «Le due figure viste mi sembrano sovrapposte in *questo* istante», oppure rispondere «*Sì*, il mio giudizio immediato, affrettato (poiché mi hai detto che la velocità è essenziale) è che la parola che ho appena sentito era compresa nella lista che hai letto poco fa». Per molti scopi sperimentali, allora, vorremo esplicitare il significato di queste pressioni di pulsanti e incorporarle nel testo come elementi a pieno titolo. Quale atto linguistico sarà rappresentato dalla pressione di un particolare pulsante dipenderà dalla interpretazione intenzionale delle interazioni tra soggetto e sperimentatore avvenute nella preparazione del soggetto all'esperimento. (Non tutte le pressioni di un pulsante equivalgono ad atti linguistici; alcuni possono, per esempio, simulare la guida di un razzo o lo sparare con una pistola.)

Quando sorgono dei dubbi sul fatto che il soggetto abbia effettivamente detto quello che intendeva o capito il problema o conosca il significato delle parole usate, possiamo chiedere delle chiarificazioni. Di solito possiamo eliminare i dubbi. Idealmente, l'effetto di queste precauzioni è di rimuovere tutte le possibili fonti di ambiguità e incertezze dalla situazione sperimentale, così che *una* interpretazione intenzionale del testo (incluse le pressioni di pulsanti) non abbia rivali plausibili e la si possa considerare *la* sincera, affidabile espressione da parte di un *soggetto singolo e coerente* delle credenze e opinioni[4] di quello stesso soggetto. Come vedremo, però, talvolta questa presupposizione è problematica – specialmente quando i nostri soggetti

mostrano qualche patologia. Come trattare, per esempio, le lamentele apparentemente sincere di cecità nel caso della cosiddetta cecità isterica o gli altrettanto apparentemente sinceri dinieghi di cecità nei ciechi con anosognosia (rifiuto della cecità o sindrome di Anton)? Questi fenomeni saranno esaminati nei capitoli successivi, e ci accorgeremo che, se dobbiamo arrivare a ciò che queste persone stanno provando, non sarà solo grazie ad una diretta intervista.

4. *I mondi della narrativa e i mondi dell'eterofenomenologia*

Oltre ai particolari problemi sollevati dai casi strani, sembra esserci un problema più generale. Non è proprio questa prassi di interpretare il comportamento verbale in questo modo a presupporre la coscienza del soggetto e quindi a scavalcare l'ipotesi dello zombi? Supponiamo di trovarci di fronte ad un calcolatore «parlante» e supponiamo di riuscire ad interpretare gli output prodotti come atti linguistici che esprimono le sue credenze e opinioni, presumibilmente «riferite» ai suoi stati coscienti. Il fatto che *esista* un'interpretazione unica e coerente di una sequenza del comportamento verbale non stabilisce che tale interpretazione sia vera; potrebbe solo *sembrare* che il «soggetto» sia cosciente; rischiamo di essere ingannati da uno zombi privo di qualsiasi vita interiore. Con questo metodo di interpretazione non potremmo *confermare* che il calcolatore sia cosciente di qualcosa. Sono d'accordo. Non possiamo essere sicuri che gli atti linguistici osservati esprimano reali credenze su effettive esperienze; forse esprimono solo *apparenti* credenze su esperienze *non esistenti*. Eppure, trovare anche una sola interpretazione stabile del comportamento di qualche entità come atti linguistici sarebbe sempre un fatto degno di attenzione. Chiunque trovasse un modo intersoggettivamente uniforme di interpretare l'ondeggiamento dei rami di un albero nel vento come «un'analisi» da parte del «clima» degli attuali eventi politici avrebbe trovato qualcosa di meraviglioso e che richiede una spiegazione, anche se si rivelerà essere l'effetto di un ingegnoso marchingegno creato da qualche scienziato in vena di burle.

Fortunatamente, abbiamo a disposizione un'analogia che ci aiuta a *descrivere* tali fatti senza dover nello stesso tempo *spiegarli* presuntivamente: possiamo paragonare il compito dell'eterofenomenologo di interpretare il comportamento dei soggetti con quello del lettore di interpretare un'opera narrativa. Alcuni

testi, come i romanzi o i racconti, sono – o si presuppone che siano – fittizi, ma ciò non impedisce una loro interpretazione. In realtà, ciò rende per alcuni aspetti più facile il compito dell'interpretazione, poiché elimina o pospone le questioni difficili sulla sincerità, la verità e i riferimenti.

Consideriamo alcuni fatti non controversi sulla semantica del racconto (Walton, 1973, 1978; Lewis, 1978; Howell, 1979). Un romanzo racconta una storia, ma non una storia vera, tranne che per caso. Pur sapendo o presupponendo che la storia non sia vera, noi possiamo parlare di ciò che *è vero nella storia* (e lo facciamo). «Possiamo dire veridicamente che Sherlock Holmes viveva in Baker Street e che gli piaceva mettere in mostra le sue capacità intellettuali. Non possiamo dire altrettanto veridicamente che era un devoto uomo di famiglia o che lavorasse in stretta collaborazione con la polizia» (Lewis, 1978, p. 37). Ciò che è vero nella storia è molto, molto di più di ciò che viene esplicitamente asserito nel testo. È vero che non c'erano aeroplani a reazione nella Londra di Holmes (sebbene ciò non sia esplicitamente asserito o nemmeno logicamente implicato nel testo), ma è anche vero che c'erano accordatori di pianoforti (sebbene – per quanto mi ricordi – non ne è menzionato nessuno o, di nuovo, logicamente implicato). Oltre a quello che è vero o falso, nella storia c'è anche una grande massa di fatti indeterminati: malgrado sia vero che Holmes e Watson presero il treno alle 11.10 dalla stazione di Waterloo per Aldershot un giorno d'estate, non è né vero né falso che quel giorno fosse un mercoledì (*The Crooked Man*).

Esistono degli squisiti problemi filosofici su come dire (in senso stretto) tutto ciò che vogliamo dire con sicurezza quando parliamo di finzione narrativa, ma essi non ci riguardano. Forse alcune persone hanno profonde perplessità sullo stato metafisico della gente e degli oggetti della narrativa, ma io no. Nel mio allegro ottimismo non credo che ci sia nessun profondo problema filosofico sul modo in cui dobbiamo rispondere, ontologicamente, ai prodotti della narrativa; la finzione narrativa è *finzione*; non esiste *nessun* Sherlock Holmes. Mettendo da parte, allora, queste complicazioni e le ingegnose proposte tecniche per trattarle, io voglio attirare l'attenzione su di un semplice fatto: l'interpretazione della finzione narrativa è sicuramente fattibile, e con alcuni risultati incontrovertibili. In primo luogo, la specificazione della storia, l'esplorazione del «mondo di Sherlock Holmes», per esempio, non è oziosa e inutile; si possono imparare molte cose sul romanzo, sul testo, sull'autore,

sugli scopi, perfino sul mondo reale, imparando dal *mondo rappresentato* nel romanzo. Secondo, se evitiamo identificazioni affrettate e giudizi personali o partigiani (per esempio, «Watson è un noioso pedante»), possiamo accumulare un'enormità di fatti oggettivi e incontrovertibili sul mondo rappresentato. Tutti gli interpreti concordano nel dire che Holmes era più intelligente di Watson; l'oggettività risiede nelle palesi ovvietà.

Terzo – e questo fatto dà un grande sollievo agli studenti – la conoscenza del mondo rappresentato da un romanzo può essere indipendente dalla conoscenza dell'effettivo testo del romanzo. Io potrei probabilmente scrivere una soddisfacente tesina su *Madame Bovary*, sebbene non abbia mai letto il libro – nemmeno nella traduzione inglese. Ho visto la serie televisiva della BBC e quindi conosco la storia; so cosa accade in quel mondo. Quello che voglio dire è questo: i fatti che riguardano il mondo di una storia fittizia sono fatti ad un livello puramente *semantico* di quella storia; essi sono indipendenti dai fatti sintattici del testo (se la narrazione è un testo). Possiamo paragonare il musical o il film *West Side Story* con la tragedia di Shakespeare *Romeo e Giulietta*; descrivendo le somiglianze e le differenze tra questi mondi, noi comprendiamo alcune somiglianze nelle opere d'arte che non sono descrivibili nei termini specifici della descrizione sintattica o testuale (per non parlare di quella fisica) delle concrete realizzazioni delle storie fittizie. Il fatto che in ogni mondo ci sia una coppia di amanti che appartengono a fazioni differenti non è un fatto che riguarda il vocabolario, la struttura delle frasi, la lunghezza (in parole o in fotogrammi), oppure forma, dimensione e peso di qualunque particolare supporto fisico di queste opere.

In generale, si può descrivere che cosa è rappresentato in un'opera d'arte (per es., *Madame Bovary*) senza descrivere *come* tale rappresentazione è realizzata. (Naturalmente, questa separazione non viene di solito fatta, e si mescolano commenti sul mondo rappresentato a commenti sugli strumenti utilizzati dall'autore per rappresentarlo, ma la separazione è possibile.) Si può persino immaginare di conoscere a sufficienza il mondo rappresentato da essere in grado di identificare l'autore della storia, pur ignorando il testo o qualsiasi altra fedele trascrizione. Si può apprendere indirettamente che cosa accade nella storia ed esclamare: solo Wodehouse poteva aver inventato una disavventura così assurda. Pensiamo di poter identificare dei generi di eventi e circostanze (e non solamente dei generi di *descrizioni* di eventi e circostanze) come kafkiane o siamo pronti

a dichiarare dei caratteri come squisitamente shakespeariani. Molte di queste plausibili convinzioni sono senza dubbio erronee (come potrebbe essere mostrato da ingegnosi esperimenti), ma non tutte. Io vi ho accennato solamente per illustrare quanto si possa «spigolare» da *ciò che è rappresentato*, malgrado non si abbia una grande conoscenza su *come la rappresentazione* è realizzata.

Applichiamo ora l'analogia al problema che si trova di fronte lo sperimentatore che vuole interpretare i testi prodotti dai soggetti senza scartare l'ipotesi che essi possano essere zombi, calcolatori, bugiardi o confusi. Consideriamo i vantaggi che si ottengono adottando la tattica di interpretare questi testi come specie di narrazioni fittizie, non come letteratura naturalmente, ma in grado di produrre una *finzione dello sperimentatore* (che dopo tutto si può dimostrare vera, naturalmente). Il lettore di un romanzo lascia che il testo *costituisca* un mondo (fittizio), un mondo che sgorga dal testo, ricco e particolareggiato fin tanto che l'estrapolazione è possibile, indeterminato oltre tal punto; il nostro sperimentatore, l'eterofenomenologo, lascia che il testo di un soggetto *costituisca il mondo eterofenomenologico* di quel soggetto, un mondo che sgorga dal testo (interpretato) e indeterminato per il resto. Ciò permette all'eterofenomenologo di posporre i problemi spinosi sulle plausibili relazioni tra quel mondo (fittizio) e il mondo reale. Ciò permette agli scienziati di trovare un accordo dettagliato sulle caratteristiche del mondo eterofenomenologico del soggetto, pur presentando delle descrizioni completamente differenti su come quel mondo eterofenomenologico si associa a eventi nel suo cervello (o nella sua anima). Il mondo eterofenomenologico del soggetto sarà un postulato teorico stabile e intersoggettivamente confermabile, e avrà lo stesso status metafisico, per esempio, della Londra di Sherlock Holmes o del mondo secondo Garp.

Come nella narrativa, quello che l'autore (l'autore apparente) dice è legge. Più precisamente, quello che l'autore apparente dice fornisce un testo che, una volta interpretato secondo le regole appena menzionate, stabilisce la maniera in cui un certo «mondo» *è*. Noi non chiediamo come Conan Doyle viene a sapere il colore della poltrona di Holmes e non solleviamo la possibilità che si possa essere sbagliato; correggiamo però gli errori tipografici e cerchiamo la lezione migliore e più coerente del testo. Allo stesso modo, non chiediamo come i soggetti (i soggetti apparenti) sanno quello che asseriscono e non ci sogniamo neanche di pensare (almeno in questo

momento) che possano sbagliarsi; ci fidiamo della loro parola (interpretata). Nota, inoltre, che malgrado i romanzi spesso includano un'avvertenza che mette in guardia il lettore dall'interpretare i fatti descritti come relativi a persone reali, vive o morte, la tattica di lasciare che il testo costituisca un mondo non deve essere limitata alle opere letterarie *volutamente progettate* come finzione dai loro autori; noi possiamo descrivere la Regina Vittoria di un certo biografo o il mondo di Henry Kissinger, infischiandocene allegramente delle presunte intenzioni dell'autore di raccontare la verità su persone reali.

5. *Il fascino discreto dell'antropologo*

Questo modo di trattare le persone come produttori di finzioni narrative (dello sperimentatore) non è il nostro modo normale di trattare le persone. *Concedere* semplicemente autorità costitutiva alle loro affermazioni può sembrare piuttosto condiscendente, e piuttosto che mostrare un rispetto autentico ne offre uno falso. Ciò emerge chiaramente quando questa tattica eterofenomenologica viene adottata dagli antropologi. Un esempio renderà immediatamente chiaro ciò che intendo dire. Supponiamo che alcuni antropologi scoprano una tribù che crede in una divinità della foresta finora sconosciuta, il dio Phaenomeno. Gli antropologi si trovano di fronte a una scelta radicale: possono convertirsi alla religione degli indigeni e credere sinceramente alla reale esistenza e al benevolo operare di Phaenomeno; oppure possono *studiare il culto* con un atteggiamento agnostico. Consideriamo questa seconda possibilità. Pur non credendo in Phaenomeno, gli antropologi decidono di studiare e sistematizzare, nel miglior modo possibile, la religione di questa tribù. Seguendo le informazioni fornite dagli indigeni, abbozzano descrizioni di Phaenomeno; cercano concordanze, ma non sempre le trovano (qualcuno dice che Phaenomeno ha gli occhi azzurri, altri che Lui – o Lei – ha gli occhi castani). Tentano di spiegare e di eliminare le discordanze, identificando e ignorando i furboni, esaminando le riformulazioni dei loro informatori e forse perfino mediando nelle dispute. A poco a poco, emerge una costruzione logica: Phaenomeno, il dio della foresta, con un elenco completo di attributi e abitudini, e una biografia. Questi scienziati agnostici (che si autodefiniscono phaenomenologi) hanno descritto, ordinato e catalogato una parte del mondo costituita dalle credenze degli indigeni, e (se

hanno eseguito bene il loro lavoro di interpretazione) hanno compilato la descrizione *definitiva* di Phaenomeno. Le credenze dei credenti indigeni (potremmo chiamarli phaenomenisti) sono autorevoli (è il *loro* dio, dopo tutto), ma solo perché Phaenomeno è stato trattato *meramente* come un «oggetto intenzionale», una mera finzione per quel che riguarda i miscredenti, e quindi interamente *creato* dalle credenze (vere o false) dei phaenomenisti. Dal momento che quelle credenze possono essere in contrasto, Phaenomeno, come costrutto logico, può vedersi attribuire proprietà contraddittorie – ma questo non è un problema per i phaenomenologi poiché per loro è *solo* un costrutto logico. I phaenomenologi tentano di presentare il costrutto logico migliore, ma non sono tenuti a risolvere tutte le contraddizioni; anzi, sono preparati a scoprire, tra i devoti, divergenze non risolte e non eliminabili.

I phaenomenisti, naturalmente, non la pensano allo stesso modo – per definizione, poiché essi sono i credenti, per i quali Phaenomeno non è un *mero* oggetto intenzionale, ma qualcuno tanto reale quanto te o me. Riguardo alla loro autorevolezza sugli attributi di Phaenomeno, il loro atteggiamento è – o dovrebbe essere – un po' più complicato. Da una parte, essi sono convinti di *sapere* tutto su Phaenomeno – sono phaenomenisti, dopo tutto: chi può conoscerlo meglio di loro? Eppure, a meno che non si ritengano dotati di una sorta di infallibilità di tipo papale, dovrebbero ammettere che in linea di principio possano sbagliarsi su qualche dettaglio. È possibile che arrivino ad essere istruiti sulla vera natura di Phaenomeno; per esempio, lo stesso Phaenomeno potrebbe chiarire loro alcuni particolari. Così, essi si potrebbero trovare perfino a disagio di fronte alla esagerata credulità (come appare loro) dei ricercatori phaenomenologi, che quasi sempre accettano scrupolosamente le loro parole, senza mai controbattere o dubitare, cercando solo di risolvere con altre domande le ambiguità e le più manifeste contraddizioni. Un phaenomenista indigeno che si trovasse d'accordo con i visitatori phaenomenologi e si schierasse sulle loro posizioni dovrebbe prendere le distanze dalle proprie convinzioni (o sarebbe meglio dire dalle proprie *precedenti* convinzioni?) e perciò dovrebbe uscire dalle file dei veri devoti.

Il metodo eterofenomenologico né respinge né accetta come interamente vere le affermazioni dei soggetti, ma mantiene piuttosto una neutralità costruttiva e comprensiva, nella speranza di compilare una descrizione *definitiva* del mondo così come appare ai soggetti. Ma loro potrebbero risentirsi per

questa concessione di semplice autorità costitutiva e protestare: «No, *davvero*! Ciò che sto descrivendoti è *perfettamente reale* ed ha esattamente le proprietà che ti sto dicendo!»: L'eterofenomenologo potrebbe rispondere onestamente con un cenno d'assenso e rassicurare il soggetto che ovviamente la sua sincerità non è messa in dubbio. Ma poiché i credenti di solito vogliono di più – vogliono che alle loro affermazioni si creda e, se ciò non avviene, vogliono sapere quando e perché i loro ascoltatori non gli credono – è in generale più diplomatico per gli eterofenomenologi, sia antropologi sia sperimentatori che studiano la coscienza in laboratorio, evitare di attirare l'attenzione sulla loro ufficiale neutralità.

Questa deviazione dalle normali relazioni interpersonali è il prezzo che deve essere pagato per ottenere la neutralità necessaria a una scienza della coscienza. Ufficialmente dobbiamo tenere aperta la possibilità che i nostri apparenti soggetti siano dei bugiardi, degli zombi o dei pappagalli vestiti da persone, ma non dobbiamo rischiare di innervosirli troppo propagandando questo fatto. Per di più, questa tattica neutrale è solo un punto di passaggio sulla strada che conduce alla formulazione e alla conferma di una teoria empirica che potrebbe in linea di principio rendere giustizia ai soggetti.

6. *Scoprire di cosa si parla realmente*

Come si potrebbero confermare le credenze di una persona sulla sua fenomenologia? Cerchiamo di capirlo con l'aiuto delle nostre analogie. Vediamo come potremmo confermare che un qualche «romanzo» è in realtà una biografia vera (o sostanzialmente vera). Potremmo cominciare chiedendoci: su quale persona reale conosciuta dall'autore è modellato questo personaggio? Questo personaggio non è la madre dell'autore camuffata? Quali eventi dell'infanzia dell'autore sono stati trasformati in questo episodio inventato? Cosa sta *realmente* cercando di dire l'autore? Chiedere all'autore non è sempre il modo migliore per rispondere a queste domande, poiché egli potrebbe non saperlo. Talvolta si può arguire plausibilmente che l'autore è stato forzato, involontariamente, a esprimersi in termini allegorici o metaforici; le uniche risorse espressive che aveva a disposizione non gli hanno permesso – per qualunque ragione – di fornire un resoconto diretto, fattuale e non metaforico degli eventi che voleva raccontare; la storia che ha composto si può considerare

un compromesso, un residuo. E come tale può essere reinterpretato drasticamente (se necessario malgrado le ripetute proteste dell'autore) e affermare che rivela un racconto vero, su persone ed eventi reali. Per esempio si potrebbe dire: poiché non è certo un caso che questo personaggio fittizio abbia questi tratti, potremmo reinterpretare il testo che rappresenta questo personaggio in modo tale che si possa vedere un riferimento – genuino, non fittizio – tra le sue parole e i tratti e le azioni di una persona reale. Rappresentare il personaggio fittizio di Molly come una sgualdrina potrebbe essere giustamente considerato come un modo per calunniare la vera Polly, giacché tutti i discorsi su Molly sono *in realtà* su Polly. Le proteste dell'autore contro tale identificazione ci possono convincere, giustamente o ingiustamente, che la calunnia non era, in ogni caso, deliberata o cosciente, ma siamo stati da troppo tempo persuasi da Freud e da altri che gli autori, come tutti noi, sono spesso all'oscuro sulle fonti più recondite delle loro intenzioni. Se ci può essere una calunnia inconscia, ci deve essere anche un riferimento involontario ad essa associato.

O, per ritornare all'altra analogia, consideriamo cosa succederebbe se un antropologo confermasse che esiste davvero un tizio dagli occhi azzurri chiamato Phaenomeno che guarisce gli ammalati e svolazza nella foresta come Tarzan. Non un dio, e non capace di volare o di essere in due posti contemporaneamente, ma tuttavia indubbiamente la fonte reale della maggior parte degli avvistamenti, leggende e credenze dei phaenomenisti. Ciò sicuramente provocherebbe laceranti disillusioni tra i devoti; alcuni dei quali si adatteranno ad una revisione e diminuzione del loro credo, altri rimarranno fermi alla versione ortodossa, pur dovendo associare forzosamente il «vero» Phaenomeno (con i suoi poteri sovrannaturali intatti) al suo agente in carne ed ossa nel mondo. Si può benissimo comprendere la resistenza degli ortodossi all'idea che si erano sbagliati così tanto su Phaenomeno; e a meno che il candidato proposto dagli antropologi quale reale referente della dottrina phaenomenista non presenti una decisa rassomiglianza, in proprietà ed opere, al Phaenomeno costituito dalla leggenda, esso non sarà neanche preso in considerazione. (Confronta questa situazione con l'affermazione: «Ho scoperto che Babbo Natale esiste davvero. È un violinista, alto e magro, che vive a Miami sotto il nome di Fred Dudley; odia i bambini e non compra mai alcun regalo».)

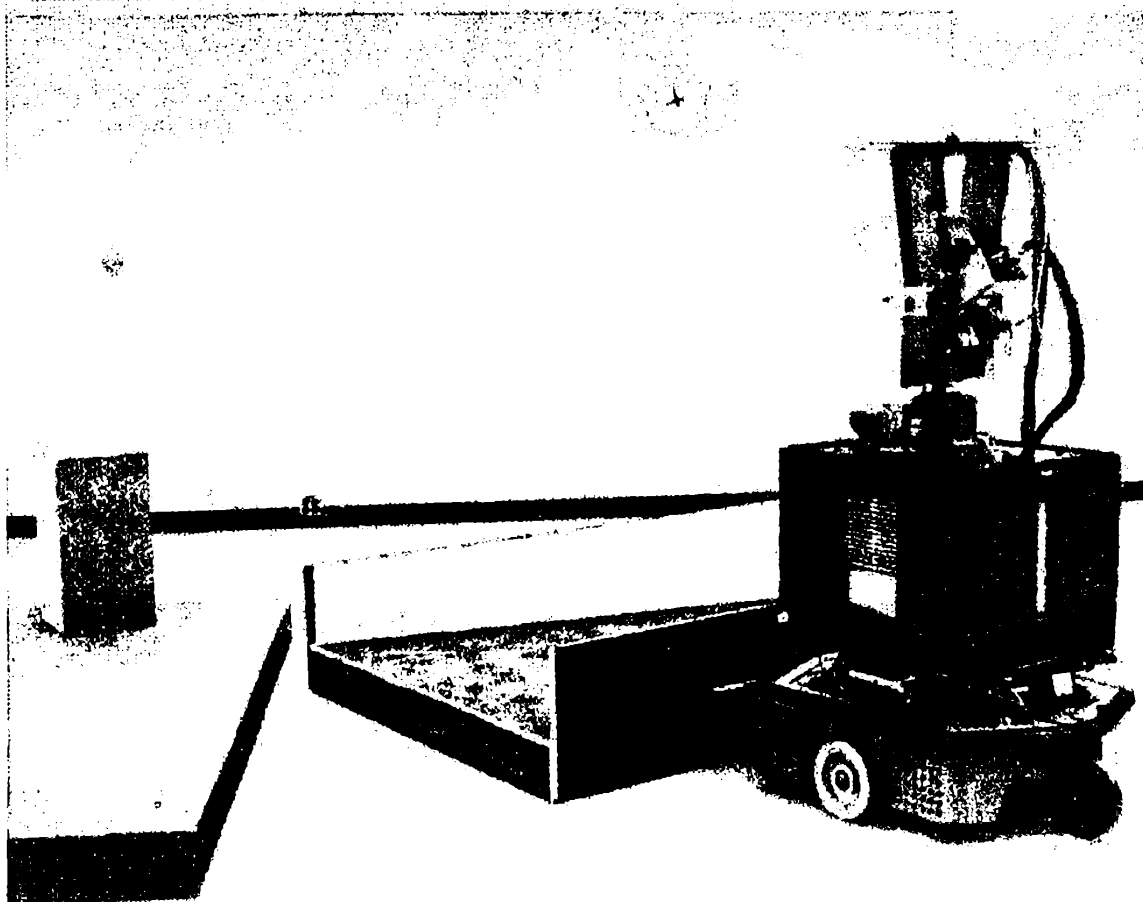
Io suggerisco, allora, che potremmo ragionevolmente proporre di aver scoperto ciò di cui una persona sta *realmente*

parlando (malgrado la sua iniziale resistenza all'identificazione), se trovassimo dei reali avvenimenti nel cervello di tale persona che avessero *sufficienti* caratteristiche in comune con gli oggetti che popolano il suo mondo eterofenomenologico. E se scoprissimo che questi avvenimenti reali hanno soltanto una rassomiglianza minima con gli oggetti eterofenomenologici, potremmo ragionevolmente dichiarare che tale persona, malgrado la sua sincerità, stava credendo qualcosa di errato. Sarebbe sempre possibile, comunque, che qualcuno – come l'irriducibile phaenomenista – insista nel dire che i reali oggetti fenomenologici *accompagnano* gli avvenimenti cerebrali senza essere identici ad essi, ma ci sarebbe poi da chiedere se una tale pretesa sarebbe anche minimamente convincente.

Come gli antropologi, possiamo rimanere neutrali mentre indaghiamo il problema. Questa neutralità potrebbe sembrare assurda – non è semplicemente inimmaginabile che degli scienziati possano scoprire alcuni fenomeni neurofisiologici che *siano esattamente* gli oggetti adorati dai soggetti nella loro eterofenomenologia? Gli eventi cerebrali sembrano troppo differenti dagli oggetti fenomenologici per essere i reali referenti delle credenze che esprimiamo nei nostri resoconti introspettivi. (Come abbiamo visto nel capitolo 2, per sostanziare la mucca viola e simili sembra necessaria una sostanza mentale.) Sospetto che la maggior parte della gente trovi ancora la prospettiva di questa identificazione del tutto inimmaginabile, ma, invece di concedere che sia davvero impossibile, io voglio provare ad allargare ulteriormente gli orizzonti della nostra immaginazione con un'altra favola. Questa riguarda un oggetto fenomenologico particolarmente enigmatico, l'*immagine mentale*, ed ha il pregio di essere in gran parte vera, sebbene semplificata e abbellita.

7. *Le immagini mentali di Shakey*

Nella breve storia dei robot, Shakey, sviluppato allo Stanford Research Institute di Menlo Park, California, alla fine degli anni Sessanta da Nils Nilsson, Bertram Raphael e i loro colleghi, merita una posizione leggendaria, non perché sapesse fare qualcosa particolarmente bene o fosse una simulazione particolarmente realistica di una caratteristica della psicologia umana, ma perché con il suo comportamento alieno ha contribuito ad aprire alcune linee di pensiero e a chiuderne altre (Raphael,

1976; Nilsson, 1984). Era il tipo di robot che un filosofo può ammirare, una specie di argomento semovente.

4.1

Shakey consisteva di uno scatolone su rotelle munito di un occhio televisivo, e invece di portarsi appresso il suo cervello vi era collegato via radio (si trattava di un ingombrante calcolatore stazionario di quei tempi). Shakey viveva in poche camere in cui gli unici altri oggetti erano alcune scatole, piramidi, rampe e piattaforme, vistosamente colorate per facilitare la sua «visione». Si poteva comunicare con Shakey digitando dei messaggi su un terminale connesso con il suo calcolatore-cervello, in un vocabolario molto limitato del linguaggio naturale; «SPINGI LA SCATOLA FUORI DALLA PIATTAFORMA», per esempio, avrebbe sollecitato Shakey a trovare la scatola, individuare la rampa, metterla in posizione, salire sulla piattaforma sfruttando la rampa e spingere fuori la scatola.

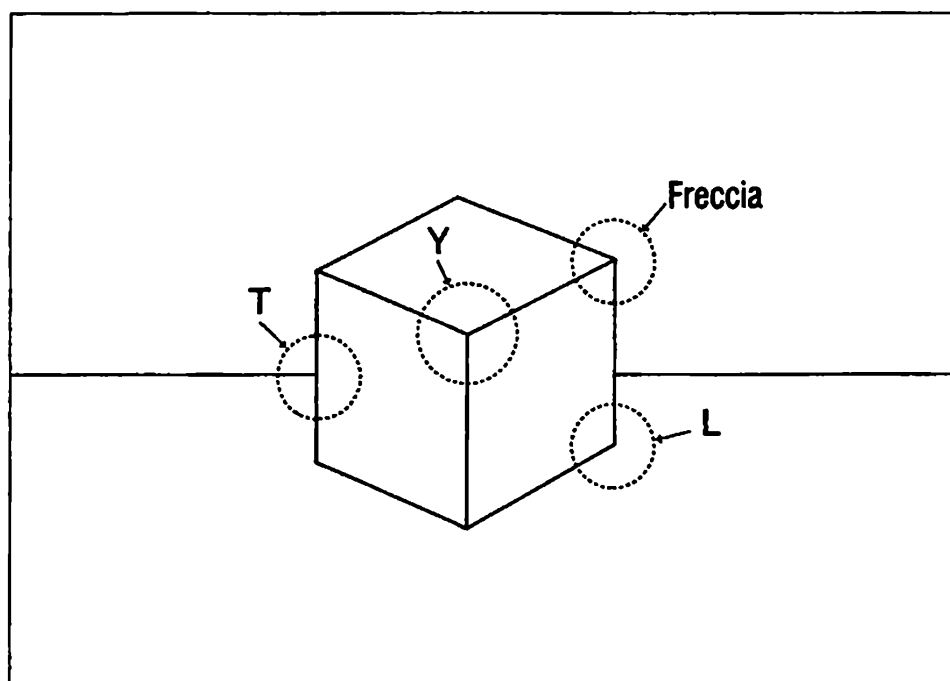
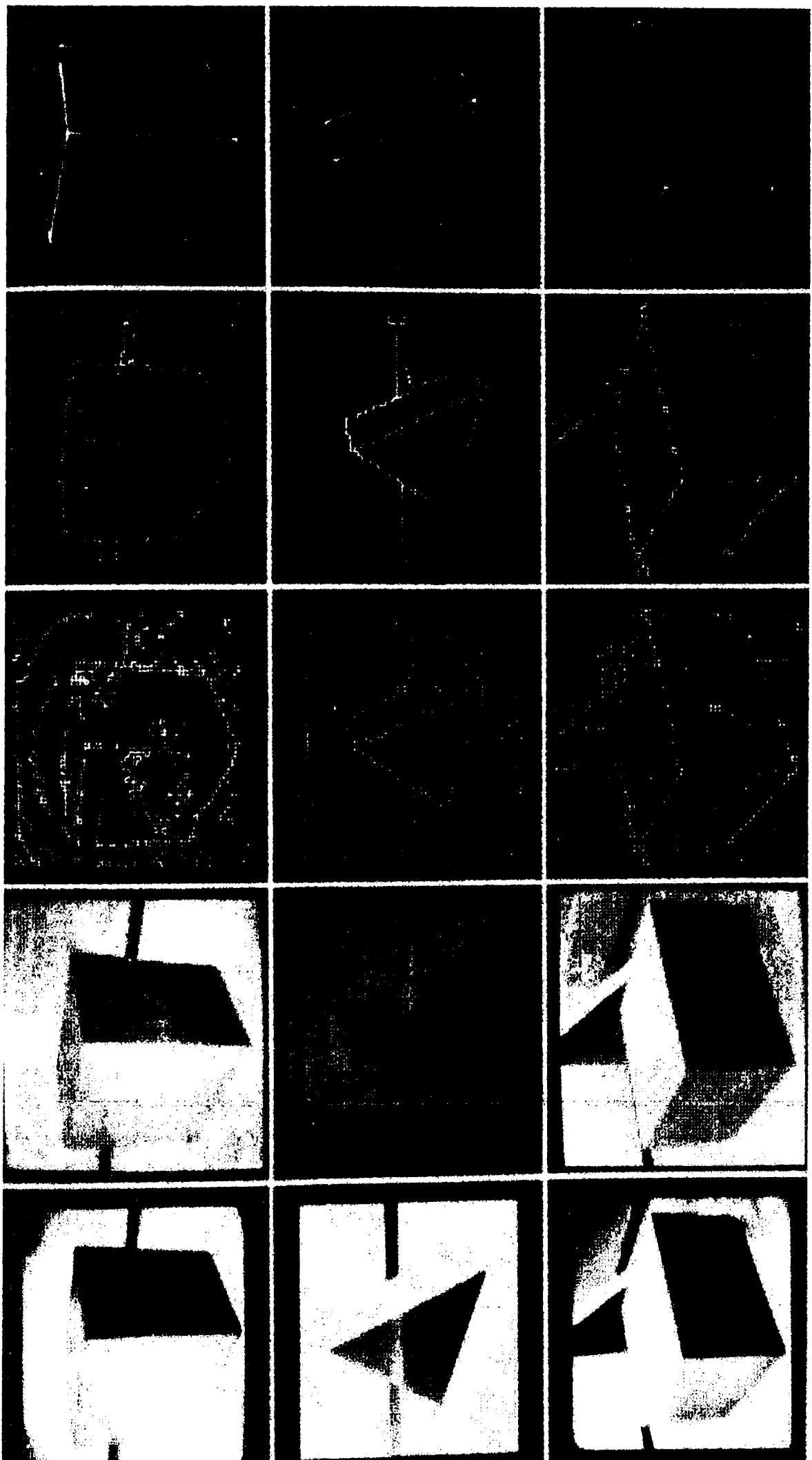
Come riusciva Shakey a fare tutto ciò? Conteneva, forse, un nanerottolo al suo interno, che osservava uno schermo televisivo e spingeva i pulsanti di controllo? Utilizzare un singolo e intelligente homunculus per realizzare l'impresa sarebbe possibile in linea di principio ma equivarrebbe a barare. Un altro

modo sarebbe quello di collegare via radio Shakey ad un controllore umano situato in un'altra stanza. Questa soluzione somiglierebbe a quella cartesiana, dove la radio rice-trasmittente gioca il ruolo della ghiandola pineale e i segnali radio sono la controparte non miracolosa dei messaggi non fisici dell'anima. La vacuità di tali «soluzioni» è evidente; ma quale *sarebbe* una soluzione non vacua? Dapprincipio può sembrare inconcepibile – o almeno incredibilmente complessa – ma sono proprio questi ostacoli all'immaginazione che noi dobbiamo fronteggiare e superare. Si rivela molto più facile del previsto immaginare come Shakey esegua i suoi compiti senza l'aiuto di un *homo ex machina.*

In particolare, come faceva Shakey a distinguere le scatole dalle piramidi con l'aiuto del suo occhio televisivo? La risposta, nelle sue linee generali, era evidente per degli osservatori esterni che potevano seguire lo svolgimento dell'intero processo in un monitor collegato al calcolatore. Dapprincipio su questo schermo sarebbe apparso un unico fotogramma a grana grossolana di una scatola, per esempio; poi l'immagine sarebbe stata purificata, rettificata e raffinata in vari modi fino a far risaltare, marcati in bianco, i precisi confini della scatola; tutta l'immagine sarebbe stata così trasformata in un disegno fatto di sole linee.

Shakey poi avrebbe analizzato il disegno: ogni vertice poteva essere identificato come una L o una T o una X o una freccia o una Y. Se fosse stato scoperto un vertice a Y, l'oggetto

4.2

4.3

doveva essere una scatola, poiché mai una piramide potrebbe presentarsi con un vertice a Y.

Questa è sicuramente una semplificazione, ma illustra i principi generali sui quali ci si basava; Shakey era dotato di un programma di «semantica delle linee» che utilizzando simili regole generali poteva determinare la categoria degli oggetti la cui immagine compariva sullo schermo. Ci si sarebbe potuto aspettare, comunque, che gli osservatori, alle prese con il monitor, soffrissero di un'improvvisa vertigine non appena avessero riflettuto sulla strana situazione in cui si trovavano: *essi* stavano osservando un processo di trasformazione di un'immagine su uno schermo, ma *Shakey non stava affatto guardandolo.* Shakey, inoltre, non stava guardando nessun altro schermo sul quale le stesse immagini venivano trasformate e analizzate. Non c'erano altri schermi nell'apparecchiatura, e inoltre lo schermo che stavano utilizzando poteva essere spento senza influenzare minimamente il processo di analisi percettiva di Shakey. Questo schermo era una sorta d'inganno? E a beneficio di chi? Solo degli osservatori. Che relazione c'era, allora, tra gli eventi che osservavano sul monitor e gli eventi all'interno di Shakey?

Il monitor era finalizzato agli osservatori, ma l'*idea* del monitor era finalizzata anche ai progettisti di Shakey. Si consideri il compito quasi inimmaginabile che si trovavano di fronte: come diamine fare a escogitare un affidabile identificatore di scatole che sfrutta soltanto dei segnali in uscita da una cinepresa televisiva? Tra le miriadi di fotogrammi che la cinepresa può inviare al calcolatore, solo una piccolissima parte rappresentano delle scatole; ogni fotogramma, poi, consiste solo di una sequenza ordinata di punti bianchi e neri o pixel, accesi o spenti, zeri e uno. Come si potrebbe scrivere un programma che identificasse tutti e solo i fotogrammi che rappresentano delle scatole? Supponiamo, per semplificare, che la retina della camera sia una griglia formata da 10.000 pixel, 100 × 100. Ogni fotogramma sarebbe allora una delle possibili sequenze di 10.000 zeri e uno. A quale configurazione di zeri ed uno corrisponderebbe l'immagine di una scatola?

Tanto per cominciare, immaginiamo di porre tutti gli zero e gli uno in una matrice che riproduca effettivamente l'immagine che si forma nella videocamera, come la matrice di pixel visibili sullo schermo. Numera i pixel in ogni riga da sinistra a destra, come le parole su di una pagina (e non come la televisione commerciale che va a zigzag). Nota, poi, che le regioni oscure sono composte in prevalenza di zeri e le regioni luminose di

tale confine può essere trasformato in una netta linea verticale tramite un accurato spostamento di zeri con uno e viceversa, in modo che qualcosa come 00011000 si presenterebbe esattamente ogni cento posizioni della sequenza.

Un confine di chiaro-scuro orizzontale è altrettanto facile da localizzare: un punto della sequenza in cui a una raffica di zero fa da eco 100, 200, 300 (ecc.) numeri dopo una raffica di uno.

I confini diagonali sono un po' più ingannevoli; il programma dovrà cercare una progressione nella sequenza. Una volta che tutti i confini sono localizzati e tracciati in bianco, il disegno è completo; può intervenire a questo punto la parte più sofisticata del programma: delle «mascherine» vengono «collocate» su segmenti delle linee in modo da identificare i vertici. Una volta che i vertici sono stati identificati, è piuttosto facile utilizzare il programma di semantica delle linee per classificare gli oggetti nell'immagine – talvolta potrà essere un compito così semplice come quello di vedere se c'è un vertice a Y.

Varie caratteristiche di questo processo sono degne di nota. Prima di tutto, ogni sottoprocesso è «stupido» e meccanico; cioè, nessuna parte del calcolatore deve capire cosa sta facendo e perché, e il modo in cui ognuno dei passi meccanici viene eseguito non è affatto misterioso. Tuttavia l'organizzazione ingegnosa di questi processi stupidi e meccanici produce un congegno che *sostituisce* un attento osservatore. (Potremmo mettere l'intero sistema visivo appena descritto in una «scatola nera» il cui compito è quello di «dire a Shakey ciò che deve sapere» su ciò che gli sta di fronte, basandosi su dei segnali in ingresso formati da fotogrammi televisivi. Così facendo avremmo dimostrato che è ragionevolmente possibile vincere la tentazione di dire che l'unico modo per realizzare un tale compito è quello di mettere un piccolo uomo nella scatola nera a guardare la Tv. Abbiamo visto ora un modo in cui questo homunculus, con un compito limitato, può essere sostituito da una macchina.)

Una volta capito come viene realizzato, possiamo comprendere che, sebbene il processo sia in forte *analogia* con quello di guardare realmente (e disegnare e cancellare) dei punti bianchi e neri sullo schermo, l'effettiva localizzazione nel calcolatore delle operazioni individuali di cambiamento da zero ad uno e viceversa non conta, purché i numeri che costituiscono i temporanei «indirizzi» delle cifre individuali codifichino l'informazione su come si susseguono i pixel. Supponiamo di spegnere il monitor. Allora, sebbene non ci sia (né sia necessaria) nessuna

effettiva immagine bidimensionale localizzabile nello spazio all'interno del calcolatore (diciamo, come «configurazione attivata nell'hardware»), le operazioni sono omomorfiche (parallele) agli eventi che stavamo osservando sul monitor. Quegli eventi erano genuinamente iconici: una superficie bidimensionale di punti di fosforo eccitato che producono una forma di una particolare grandezza, colore, localizzazione e orientamento. Così in *un* senso stretto, Shakey *non* riconosce le scatole grazie ad una serie di trasformazioni di immagini: l'ultima vera immagine nel processo era quella focalizzata sul campo ricettivo della telecamera. In un altro senso, sempre stretto ma metaforico, Shakey *riconosce* le scatole grazie ad una serie di trasformazioni di immagini – il processo appena descritto, che trasforma i confini chiaro-scuro in linee nette e poi classifica i vertici. Il fatto che questo senso pur essendo stretto sia tuttavia metaforico può essere evidenziato notando che ci sono molte proprietà che le immagini *reali* hanno e che invece mancano alle «immagini» trasformate da Shakey: esse non hanno colore né dimensioni né orientamento. (Potremmo comporre un bell'indovinello per una tale immagine: sto pensando a un'immagine che non è né più grande né più piccola della Monna Lisa, non è né a colori né in bianco e nero, e non è rivolta verso nessun punto dell'orizzonte. Che cosa è?)

È molto difficile che il processo tramite il quale Shakey ricava le informazioni sugli oggetti dalla diversità di luce nel suo ambiente sia simile a quello della visione umana e probabilmente è differente dai processi visivi di qualunque creatura. Ma questo aspetto può essere ignorato per il momento, per permetterci di comprendere una possibilità piuttosto astratta di come le immagini mentali che i soggetti umani riferiscono possano essere scoperte nel cervello. La descrizione del sistema visivo di Shakey era molto semplificata per evidenziare i punti di fondamentale interesse teoretico. Ora dobbiamo imbarcarci in una situazione da fantascienza per far emergere un altro punto: dobbiamo supporre di incrociare Shakey con un altro famoso personaggio nel campo dell'intelligenza artificiale, SHRDLU, realizzato da Terry Winograd (1972) e che manipolava dei blocchi (immaginari) e poi rispondeva a delle domande su ciò che stava facendo e perché. Le risposte di SHRDLU erano ampiamente «preconfezionate» – erano delle frasi belle e pronte o degli schemi di frasi composte da Winograd e immagazzinate nel programma. SHRDLU era stato concepito per indagare astrattamente alcuni compiti di manipolazione dell'informa-

zione che ogni interlocutore verbale deve fronteggiare, non per dare un modello realistico della produzione del linguaggio umano, e questo è lo spirito del nostro esperimento ideale. (Nel capitolo 8 prenderemo in esame dei modelli più realistici di produzione del linguaggio.) Una conversazione con la nostra nuova versione di Shakey, riprogettata per includere un repertorio più sofisticato di azioni verbali, potrebbe essere la seguente:

Perché hai spostato la rampa?
PER POTER SALIRE SULLA PIATTAFORMA.
E perché vuoi salirci?
PER SPINGERE VIA LA SCATOLA.
E perché vuoi farlo?
PERCHÉ ME LO HAI CHIESTO.

Ma supponiamo poi di chiedere a Shakey:

Come distingui le scatole dalle piramidi?

Quale dovrebbe essere la risposta che facciamo «dire» a Shakey? Ecco tre possibilità:

1. Analizzo ogni sequenza di 10.000 cifre di 0 e 1 che mi fornisce la mia telecamera, cerco certe regolarità nelle sequenze, come per esempio bla,bla,bla (una risposta *molto* lunga se vogliamo che Shakey entri nei minimi dettagli).

2. Cerco i confini tra il chiaro e lo scuro e li evidenzio con delle linee bianche nel mio occhio mentale; poi guardo i vertici; se trovo un vertice a Y, per esempio, so di aver a che fare con una scatola.

3. Non lo so; alcune cose mi appaiono semplicemente a forma di scatola. Mi capita di saperle riconoscere; è grazie alla mia *intuizione*.

Qual è la risposta giusta da far dire a Shakey? Ognuna è giusta a suo modo; sono tutte descrizioni di un processo elaborativo dell'informazione a differenti livelli di profondità. La decisione relativa alla risposta da far dire a Shakey dipende ampiamente dalla decisione relativa al grado di accesso che la capacità espressiva di Shakey (la sua scatola nera SHRDLU)

avrà nei confronti dei suoi processi percettivi. Forse ci saranno buone ragioni ingegneristiche per negare un accesso profondo (dettagliato e divoratore di tempo) ai processi d'analisi intermedi. Ma quali che siano le capacità autodescrittive di cui dotiamo Shakey, ci sarà un limite alla profondità e al dettaglio della sua «conoscenza» esprimibile su ciò che sta accadendo in lui, che lui sta facendo. Se la risposta migliore che Shakey può dare è la 3, allora egli si trova, rispetto alla domanda su come distingue le scatole dalle piramidi, nella stessa posizione in cui ci troviamo noi quando ci chiedono come facciamo a distinguere la parola «sì» dalla parola «sci»; non lo sappiamo; una suona come «sì» e l'altra come «sci» – questo è il massimo che possiamo dire. E se Shakey è progettato per rispondere con la 2, allora ci saranno ancora altre domande alle quali non potrà rispondere, ad esempio, «Come tracci le linee bianche sulle tue immagini mentali?» o «Come identifichi un vertice a freccia?».

Supponiamo di progettare Shakey in modo che abbia il tipo 2 di accesso ai suoi processi di analisi percettiva; quando gli chiediamo come fa, egli ci racconta il modo in cui trasforma le immagini. Se a sua insaputa spegniamo il monitor, siamo autorizzati a dirgli che noi ne sappiamo di più? Che, in realtà, non sta elaborando delle immagini, sebbene lo creda? (Egli *dice* che lo sta facendo, e quindi, seguendo la strategia eterofenomenologica, noi interpretiamo le sue affermazioni come espressione delle sue credenze.) Se egli fosse una simulazione realistica di una persona, ci potrebbe benissimo replicare che non siamo nella posizione di poter dire *a lui* ciò che sta accadendo nella sua mente! *Egli* sa quello che sta facendo, che cosa sta *realmente* facendo! Se fosse più sofisticato, potrebbe concedere che ciò che sta facendo può essere descritto come elaborazione di immagini solo metaforicamente – sebbene egli senta una forte propensione a descrivere così ciò che sta accadendo. In tal caso noi saremmo in grado di dirgli che la sua metafora è perfettamente scelta.

Se fossimo più diabolici, d'altra parte, potremmo farci beffe di Shakey e fargli dire delle cose assurde sul modo in cui elabora le informazioni. Potremmo progettarlo in modo che desideri dire delle cose che non hanno il minimo rapporto con ciò che sta realmente accadendo («Uso le informazioni che mi giungono dalla telecamera per dirigere un cesello interiore che scolpisce in tre dimensioni un pezzo di creta mentale. Poi, se il mio homunculus può sedercisi sopra, è una scatola; se cade, è una piramide»). Non ci sarebbe nessuna interpretazione che garan-

tisca anche un minimo di verità a questa descrizione; Shakey sta semplicemente inventando una storia senza rendersene conto – produce *confabulazioni.*

E questa possibilità, in noi, ci mostra perché dobbiamo utilizzare il fastidioso accorgimento di trattare l'eterofenomenologia in modo analogo alla interpretazione della narrativa. Come abbiamo già visto, le persone in alcune circostanze si sbagliano su ciò che stanno facendo e su perché lo stanno facendo. Non è esatto dire che stanno *mentendo* nella situazione sperimentale, è vero piuttosto che si lasciano andare alla confabulazione; riempiono i vuoti, immaginano, ipotizzano, scambiano il teorizzare con l'osservare. La relazione tra ciò che dicono e ciò che li spinge a dire ciò che dicono non potrebbe essere più oscura, sia a noi eterofenomenologi all'*esterno* sia ai soggetti stessi. *Essi* non hanno nessuno strumento per «vedere» (presumibilmente, con un occhio interiore) il processo che governa le loro asserzioni, ma ciò non impedisce loro di avere delle opinioni granitiche da esprimere.

Per riassumere, le persone creano involontariamente delle storie fittizie, ma dire che lo fanno involontariamente significa concedere che ciò che dicono è, o può essere, una *descrizione esatta di come sembra a loro.* I soggetti ci dicono *cosa si prova* a risolvere un problema, a prendere una decisione, a riconoscere un oggetto. Dal momento che sono (apparentemente) sinceri, concediamo che deve esser così che sembra a loro, ma ne consegue che ciò che sembra loro è al massimo una guida incerta per capire cosa accade in loro. Talvolta può accadere che le storie che involontariamente creiamo siano vere, dopo tutto, se accettiamo una qualche imprecisione metaforica, come per la risposta 2 di Shakey. Per esempio, recenti ricerche sull'immaginazione eseguite da psicologi cognitivi mostrano che le nostre affermazioni introspettive sulle immagini mentali che riusciamo a produrre (sia di mucche viola che di piramidi) non sono del tutto false (Shepard e Cooper, 1982; Kosslyn, 1980; Kosslyn, Holtzmann, Gazzaniga e Farah, 1985). Ciò sarà discusso più dettagliatamente nel capitolo 10, dove vedremo anche come le nostre descrizioni introspettive dell'immaginazione *possano* essere interpretate in modo da renderle vere. Come il nostro Tarzan-Phaenomeno, comunque, che non si dimostra capace di volare o di essere in due luoghi contemporaneamente, le cose che effettivamente troviamo nel cervello per *identificare* le immagini mentali non avranno tutte le proprietà meravigliose che i soggetti hanno fiduciosamente attribuito loro. Le «imma-

gini» di Shakey ci forniscono un esempio di qualcosa che in realtà non è affatto un'immagine, ma che può essere proprio quella cosa di cui si parla sotto forma di immagine. Sebbene i processi cerebrali che sottostanno all'immaginazione umana probabilmente non sono affatto simili a quelli di Shakey, noi abbiamo aperto il campo a delle possibilità che erano altrimenti difficilmente concepibili.

8. *La neutralità dell'eterofenomenologia*

All'inizio di questo capitolo avevo promesso di descrivere un metodo, il metodo eterofenomenologico, che fosse neutrale rispetto all'opposizione tra approccio oggettivo e approccio soggettivo alla fenomenologia, e tra realtà fisica e realtà non fisica degli oggetti fenomenologici. Riesaminiamo il metodo per vedere se ho mantenuto la promessa.

Esaminiamo, per primo, il problema dello zombi. Molto semplicemente, l'eterofenomenologia non può di per se stessa distinguere uno zombi da una persona reale e cosciente e quindi non può pretendere né di risolvere né di liquidare il problema dello zombi. *Ex hypothesi*, gli zombi si comportano esattamente come la gente reale, e poiché l'eterofenomenologia è un modo di interpretare il comportamento (incluso il comportamento interno dei cervelli, ecc.), si arriverà esattamente allo stesso mondo eterofenomenologico sia per Silvia che per Silvia Zombi, la sua gemella non cosciente. Gli zombi hanno un mondo eterofenomenologico, ma questo significa soltanto che, quando gli sperimentatori si accingono ad interpretarli, usano esattamente gli stessi mezzi e ottengono esattamente gli stessi risultati che usiamo e otteniamo noi quando interpretiamo i nostri amici. Naturalmente, come già visto, alcuni dei nostri amici potrebbero essere zombi. (È difficile per me riuscire a restare serio su questo punto, ma, poiché alcuni filosofi molto rigorosi hanno preso il problema dello zombi molto seriamente, mi sento obbligato a fare altrettanto.)

Sicuramente non c'è nulla di sbagliato, nulla di non neutrale, nel garantire un mondo eterofenomenologico agli zombi, dal momento che si concede così poco. Questo è il minimalismo metafisico dell'eterofenomenologia. Il metodo descrive un mondo, il mondo eterofenomenologico del soggetto, in cui si trovano vari *oggetti* (gli oggetti intenzionali, nel gergo filosofico) e in cui varie cose accadono a questi oggetti. Se qualcuno

chiedesse «Cosa *sono* questi oggetti e *di cosa sono fatti?*», la risposta *potrebbe essere* «Di nulla!». Di cosa è fatto il signor Pickwick? Di nulla, il signor Pickwick è un oggetto fittizio, così come lo sono gli oggetti descritti, battezzati e menzionati dall'eterofenomenologo.

«Ma, come teorico, non è imbarazzante ammettere di parlare di entità fittizie – cose che non esistono?». Niente affatto. Gli studiosi di letteratura svolgono un lavoro onesto e prezioso descrivendo entità fittizie e lo stesso vale per gli antropologi che studiano gli dèi e le streghe di varie culture. Lo stesso fanno i fisici che quando si chiede loro di cosa è fatto un centro di gravità rispondono: «Di nulla!». Gli oggetti eterofenomenologici sono, come i centri di gravità o l'equatore, degli *abstracta* e non dei *concreta* (Dennett, 1987a, 1991a). Non sono delle oziose fantasie, ma delle utili finzioni teoriche. Inoltre, a differenza dei centri di gravità, viene lasciata aperta la possibilità di trovare dei *concreta* con cui sostituirle se il progresso della scienza empirica lo permetterà.

Esistono due modi di studiare il diluvio di Noè: si può ritenere che sia un puro e semplice mito, ma sicuramente degno di studio, oppure ci si può chiedere quale effettiva catastrofe geologica o meteorologica esso nasconda. Entrambe le linee di indagine possono essere scientifiche, ma la prima è meno congetturale. Se si vogliono fare delle congetture seguendo la seconda linea, la prima cosa da fare è quella di condurre un'attenta indagine seguendo la prima linea per raccogliere tutti i possibili suggerimenti. Similmente, se si vuole studiare come (o persino *se*) gli oggetti fenomenologici siano effettivi eventi nel cervello, la prima cosa da fare è quella di catalogare eterofenomenologicamente e con attenzione tali oggetti. Si rischia in questo modo di offendere i soggetti (così come gli antropologi che studiano Phaenomeno rischiavano di offendere i loro informatori), ma è il solo modo per evitare il conflitto di «intuizioni» che altrimenti passa per fenomenologia.

Come rispondere, tuttavia, all'obiezione che l'eterofenomenologia, basandosi sulla prospettiva in terza persona, non riesce nemmeno a toccare i *veri* problemi della coscienza? Nagel, come abbiamo visto, insiste su questo punto, e lo stesso fa il filosofo John Searle, che riferendosi esplicitamente al mio approccio ha ammonito: «Ricordatevi, in queste discussioni, di insistere sempre sulla prospettiva in prima persona. Il primo passo dell'abile trucco operazionista avviene quando cediamo all'invito di immaginare come potremmo *conoscere* cosa provano gli altri»

(Searle, 1980, p. 451). Le cose non stanno così: quando cadi tra le grinfie dell'eterofenomenologo, *ti viene garantita l'ultima parola.* Tu hai il diritto di curare, rivedere, rinnegare *ad libitum*; e fintantoché eviti la presunzione di *teorizzare* sulle cause o lo stato metafisico degli oggetti di cui parli, ti sarà concessa un'autorità costitutiva su qualunque cosa sostieni stia avvenendo nel tuo mondo eterofenomenologico. Tu sei il narratore, la tua parola è legge. Cosa potresti volere di più?

Se vuoi che noi *crediamo* a tutto quello che dici sulla tua fenomenologia, stai chiedendo non solo di essere preso seriamente ma di essere autorizzato ad una sorta di infallibilità papale; e questo è chiedere troppo. Tu *non* sei autorevole su ciò che sta avvenendo in te, ma solo su ciò che *ti sembra* stia avvenendo in te, e noi ti stiamo concedendo un'autorità totale e dittatoriale sulla descrizione di come ti sembra, di *cosa si prova ad essere al tuo posto*. E se ti lamenti dicendo che alcuni aspetti di ciò che provi sono ineffabili, noi eterofenomenologi ti garantiamo anche questo. Quali basi migliori avremmo per credere che sei incapace di descrivere qualcosa del fatto che (1) non lo descrivi e (2) confessi che non puoi? Naturalmente potresti mentire, ma ti concediamo il beneficio del dubbio. Se tu replicassi «Non sto dicendo semplicemente che *io* non posso descriverlo; sto dicendo che è indescrivibile!», noi eterofenomenologi prenderemmo nota che non puoi descriverlo *ora* e, poiché tu sei l'unico che lo può descrivere, è indescrivibile ora. In seguito, forse, sarai in grado di descriverlo, ma naturalmente a quel punto *esso* sarà qualcosa di differente, qualcosa di descrivibile.

Quando io proclamo che gli oggetti dell'eterofenomenologia sono finzioni teoriche, potresti risentirti (a molti succede, io credo) e dichiarare:

Questo è *proprio* quello che distingue gli oggetti della vera fenomenologia dagli oggetti dell'eterofenomenologia. I miei oggetti *auto*fenomenologici non sono oggetti fittizi – sono perfettamente *reali*, sebbene io non abbia la più pallida idea di che cosa siano fatti. Quando ti dico, sinceramente, che sto immaginando una mucca viola, io non sto solo producendo inconsciamente (come Shakey) una sequenza di parole che significa ciò e sia surrettiziamente preconfezionata per corrispondere vagamente a qualche analogo evento fisico che avviene nel mio cervello; io sto consciamente e deliberatamente riferendo l'esistenza di qualcosa che è *davvero lì*! Per me non è una mera finzione teorica!

Rifletti attentamente su questo discorso. Tu dici che non stai solo producendo inconsciamente una sequenza di parole. Beh, *stai* inconsciamente producendo una sequenza di parole; tu non hai la più pallida idea di come fai o di cosa è implicato nella sua produzione. Ma, insisti, non stai *solo* facendo questo; sai *perché* lo stai facendo; *comprendi* la sequenza di parole e *intendi* asserirla. Sono d'accordo con te. Ecco perché ciò che dici permette così bene di costruire un mondo eterofenomenologico. Se tu stessi solo ripetendo delle parole come un pappagallo, le probabilità che la sequenza di parole sia interpretabile in questo modo sarebbero infinitamente piccole. Sicuramente esiste una buona spiegazione di come e perché dici quello che dici, una spiegazione che descrive la differenza tra il dire solamente qualcosa e il dirlo intendendolo dire, *ma tu non sei ancora in possesso di una tale spiegazione*. Perlomeno non nella sua interezza. (Nel capitolo 8 indagheremo questo punto.) Probabilmente stai parlando di qualcosa di reale, perlomeno la maggior parte delle volte. Vediamo se riusciamo a scoprire di cosa si tratta.

Queste rassicurazioni non sono sufficienti per alcune persone. Alcune persone non vogliono affatto giocare con queste regole. Alcuni individui devotamente religiosi, per esempio, si offendono quando gli interlocutori accennano alla *possibilità* che ci sia una qualche altra vera religione. Queste persone non considerano l'agnosticismo neutrale, lo considerano un affronto, perché uno dei dogmi del loro credo recita che l'incredulità in sé è peccato. Quelli che la pensano così hanno diritto alle loro credenze e hanno diritto (se questa si può considerare la parola giusta) ai loro sentimenti offesi quando incontrano uno scettico o un agnostico, ma, a meno che non riescano a dominare l'ansietà che sentono quando apprendono che qualcuno non crede (ancora) a ciò che dicono, si autoescludono dal mondo della ricerca accademica.

In questo capitolo abbiamo sviluppato un metodo *neutrale* per indagare e descrivere la fenomenologia. Esso implica la registrazione e la purificazione dei *testi* prodotti da *soggetti* che (apparentemente) parlano; testi utilizzati per generare una finzione teorica, il *mondo eterofenomenologico* del soggetto. Questo mondo fittizio è popolato da tutti gli eventi, immagini, suoni, odori, impressioni, presentimenti e sentimenti che il soggetto in modo (apparentemente) sincero crede che esistano nel suo flusso di coscienza. Nella sua accezione più ampia equivale ad una rappresentazione neutrale ed esatta di *cosa si prova ad essere*

quel soggetto – negli stessi termini usati dal soggetto, secondo la migliore interpretazione che possiamo fornire.

Dopo aver ricavato una simile eterofenomenologia, lo studioso si può rivolgere alla questione di cosa possa giustificarne l'*esistenza* nei minimi dettagli. L'eterofenomenologia esiste – nel modo altrettanto incontrovertibile dei romanzi e delle altre storie fittizie. È indubbio che le persone credono di avere immagini mentali, dolori, esperienze percettive, ecc., e *questi fatti* – i fatti in cui le persone credono e che riferiscono quando esprimono le loro credenze – sono fenomeni che qualunque teoria scientifica della mente deve spiegare. Noi organizziamo i dati relativi a questi fenomeni in finzioni teoriche, «oggetti intenzionali» inseriti in mondi eterofenomenologici. A quel punto chiedersi se gli oggetti così rappresentati esistono come oggetti reali, come eventi e stati cerebrali – o mentali, in questo è lo stesso – è una questione che deve essere indagata empiricamente. Se verranno scoperti dei candidati plausibili, potremo identificarli come i referenti a lungo cercati delle parole dei soggetti; altrimenti, dovremo spiegare perché ai soggetti sembra che tali oggetti esistano davvero.

Adesso che le nostre presupposizioni metodologiche sono state chiarite, possiamo rivolgerci direttamente alla teoria empirica della coscienza. Inizieremo fronteggiando il problema relativo all'ordinamento e alla temporalizzazione degli oggetti nel nostro flusso di coscienza. Nel capitolo 5, presenterò un primo abbozzo della teoria e mostrerò come affronta un caso semplice. Nel capitolo 6, vedremo come ci permette di reinterpretare alcuni fenomeni molto più complicati che hanno sconcertato molti pensatori. I capitoli dal 7 al 9 svilupperanno meglio l'abbozzo iniziale della teoria, mettendo in guardia contro alcune interpretazioni fuorvianti, rispondendo ad obiezioni e mostrando meglio i suoi punti di forza.

PARTE SECONDA

UNA TEORIA EMPIRICA DELLA MENTE

5

MOLTEPLICI VERSIONI CONTRO TEATRO CARTESIANO

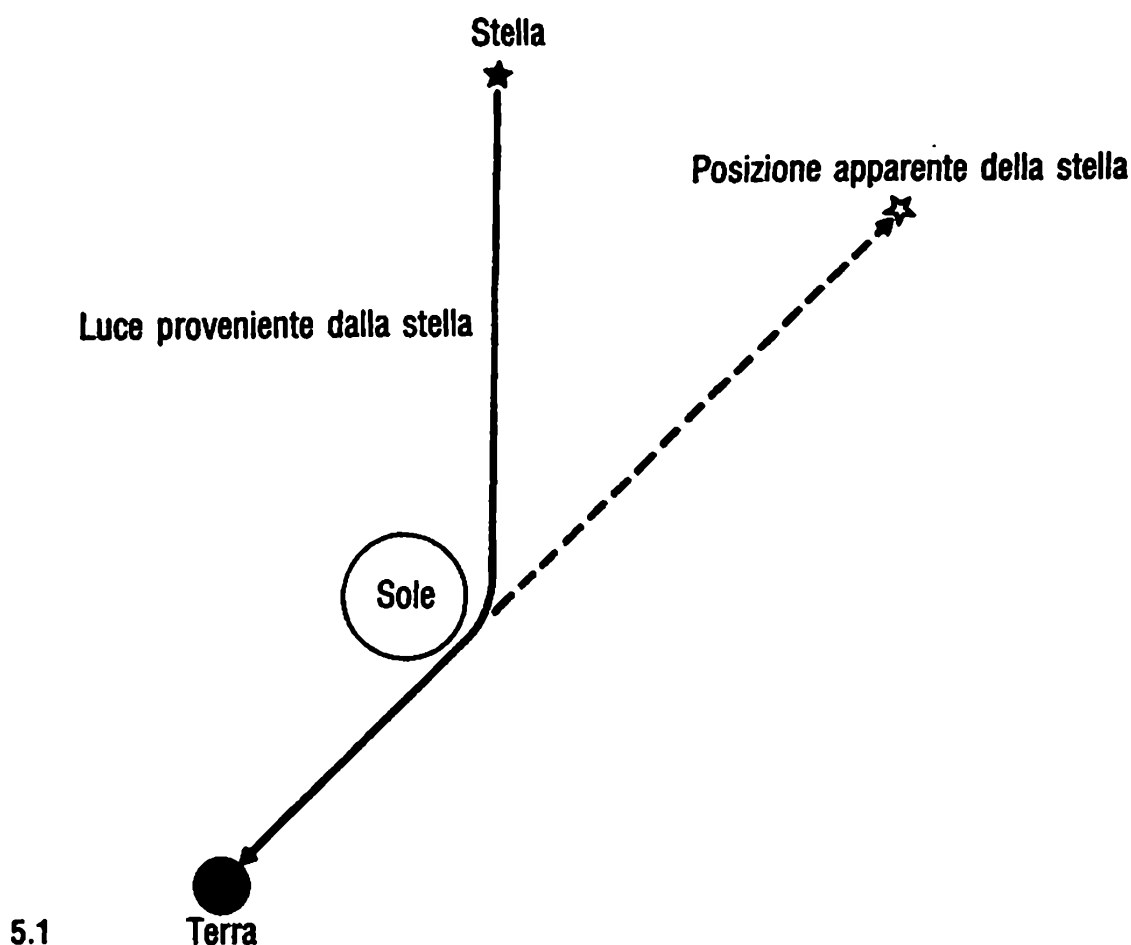
1. *Il punto di vista dell'osservatore*

> Nel cervello non esiste una cellula o un gruppo di cellule in una tale preminenza anatomica o funzionale da poter sembrare la chiave di volta o il centro di gravità dell'intero sistema.
>
> WILLIAM JAMES, 1890

I navigatori da diporto che veleggiano lungo coste insidiose di solito si dirigono verso un punto preciso per essere sicuri di evitare i pericoli; trovano una boa lontana ma ben visibile, più o meno nella direzione in cui vogliono andare, controllano sulla carta che non ci siano ostacoli nascosti sulla linea retta tra loro e il segnale, e poi puntano direttamente su di esso. Forse per un'ora o più l'unico scopo del timoniere è quello di dirigersi direttamente sul segnale e correggere le possibili deviazioni. Talvolta accade, però, che i timonieri si calano talmente in questo compito che dimenticano di virare all'ultimo minuto e colpiscono effettivamente la boa! L'obiettivo primario, stare alla larga dai pericoli, passa in secondo piano di fronte al rassicurante successo dell'obiettivo secondario, quello di puntare sul segnale. In questo capitolo vedremo come i paradossi meno spiegabili sulla coscienza sorgono dal fatto che rimaniamo troppo attaccati a una buona abitudine di pensiero, un'abitudine che di solito ci evita i problemi.

Ovunque esista una mente cosciente, esiste *un punto di vista*. Questa idea occupa un posto fondamentale nel nostro modo di concepire la mente – o la coscienza. Una mente cosciente è un osservatore, che recepisce un sottoinsieme limitato di tutte le informazioni esistenti; un osservatore che recepisce le informazioni disponibili in una sequenza particolare (più o meno) continua di tempi e luoghi nell'universo. Per la maggior parte

dei nostri scopi pratici, possiamo considerare il punto di vista di un particolare soggetto cosciente come *un semplice punto* che si muove nello spazio-tempo. Consideriamo, per esempio, gli usuali disegni che in fisica o in cosmologia illustrano l'effetto Doppler o l'effetto provocato dalla gravità sulla traiettoria della luce.

5.1

A un osservatore posto sulla superficie della Terra nella figura 5.1 corrisponde un particolare punto di vista, ad osservatori posti in differenti punti dell'universo le cose apparirebbero differenti. Esempi più semplici sono più familiari. Noi ci spieghiamo il sorprendente intervallo temporale tra l'immagine e il rumore di lontani fuochi d'artificio, chiamando in causa le diverse velocità di propagazione del suono e della luce: arrivano *all'osservatore* (a quel punto) in tempi differenti, anche se hanno lasciato la fonte nello stesso istante.

Cosa accade, però, quando focalizziamo la nostra attenzione sull'osservatore e proviamo a localizzare più precisamente il suo punto di vista, come un punto *all'interno* dell'individuo? Le semplici presupposizioni che funzionano così bene su larga scala cominciano a vacillare.[1] Non esiste un singolo punto nel cervello verso il quale tutte le informazioni vengono incanalate;

e questo fatto ha delle conseguenze tutt'altro che ovvie – anzi piuttosto controintuitive.

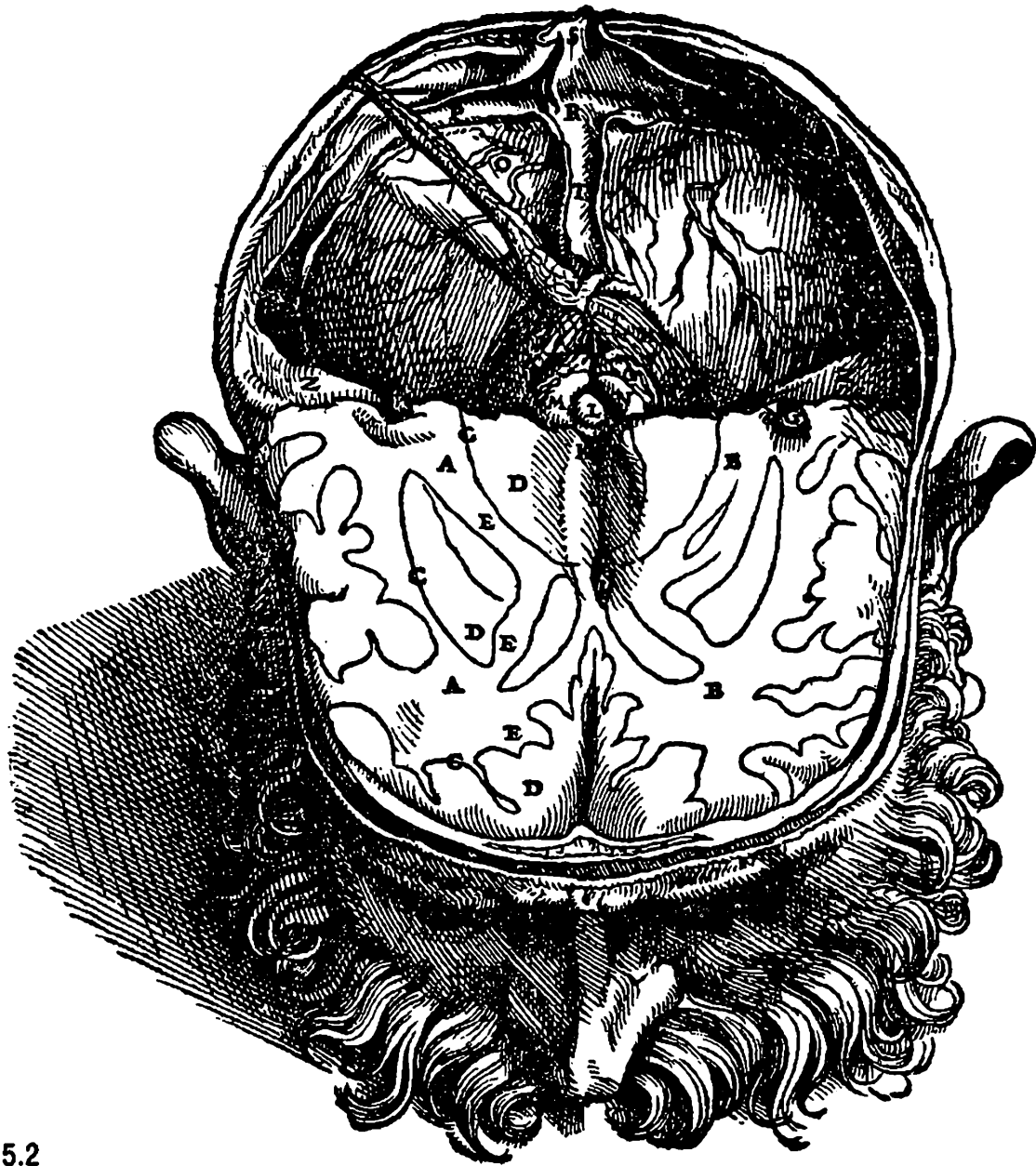
Poiché dovremo considerare degli eventi che si verificano su una scala relativamente microscopica del tempo e dello spazio, è importante avere una chiara idea delle grandezze implicate. In tutti gli esperimenti che prenderemo in considerazione gli intervalli di tempo sono misurati in millisecondi; quindi potrà essere utile sapere quanto sono lunghi (o corti) 100 o 50 msec. Puoi pronunciare circa quattro o cinque sillabe per secondo, così una sillaba prende circa 200 msec. Un film normale scorre a ventiquattro fotogrammi al secondo, così ogni fotogramma rimane esposto per 42 msec (in realtà, ogni fotogramma nel periodo che rimane stabile viene colpito dalla luce e proiettato *tre volte* durante quei 42 msec, per una durata di 8,5 msec, con intervalli di oscurità di 5,4 msec). La televisione (negli Stati Uniti) scorre a trenta fotogrammi al secondo, come dire un fotogramma ogni 33 msec (in realtà, ogni fotogramma viene passato due volte sovrapponendosi al precedente). Facendo lavorare il tuo pollice il più velocemente possibile, puoi azionare e fermare un cronometro in circa 175 msec. Quando ti dai una martellata sul dito, le fibre nervose veloci (con guaina mielinica) inviano un messaggio al cervello in circa 20 msec; le fibre C, lente e non mieliniche, inviano un segnale che impiega molto di più – circa 500 msec – per coprire la stessa distanza.

Ecco un quadro di alcuni eventi con un'approssimazione della loro durata espressa in millesecondi:

dire «mille-e-uno»	1000 msec
fibra non mielinica, dal polpastrello al cervello	500 msec
un pallone alla velocità di 110 kmh dal dischetto di rigore in porta	360 msec
pronunciare una sillaba	200 msec
azionare e fermare un cronometro	175 msec
un fotogramma di un film	42 msec
un fotogramma televisivo	33 msec
fibra veloce (mielinata), dal polpastrello al cervello	20 msec
il ciclo fondamentale di un neurone	10 msec
il ciclo fondamentale di un personal computer	0,0001 msec

Cartesio, uno dei primi pensatori a riflettere seriamente su ciò che accade all'interno del corpo dell'osservatore, ha elaborato un'idea che è così immediatamente naturale e attraente da

aver permeato sin da allora il nostro modo di pensare alla coscienza. Come abbiamo visto nel capitolo 2, Cartesio pensava che il cervello *avesse* effettivamente un centro: la ghiandola pineale, che serviva come cancello verso la mente cosciente (vedi figura 2.1). La ghiandola pineale è l'unico organo del cervello che si trovi in una posizione mediana, senza essere

5.2

duplicato in una versione destra e in una sinistra; è indicata da una «L» nel disegno del grande anatomista del sedicesimo secolo, Andrea Vesalio.

Più piccola di un pisello, risiede in splendido isolamento sul suo stelo, che la connette al resto del sistema nervoso più o meno nel centro della parte posteriore del cervello. Poiché le sue funzioni erano piuttosto oscure (e ancora lo sono), Cartesio

propose un ruolo per essa: affinché una persona sia cosciente di qualcosa, è necessario che gli impulsi provenienti dai sensi arrivino in questa stazione intermedia, dove si verifica una transazione speciale – magica, in effetti – tra il suo cervello materiale e la sua mente immateriale.

Per Cartesio, non tutte le reazioni corporee avevano bisogno dell'intervento della mente cosciente. Egli riconosceva le caratteristiche peculiari di quei movimenti che ora vengono chiamati riflessi e riteneva che fossero realizzati grazie a una sorta di corto circuito completamente meccanico che non coinvolgeva affatto la ghiandola pineale – erano quindi inconsci.

5.3

Egli si sbagliava sui dettagli: pensava che il fuoco spostasse la pelle, che tirasse un sottilissimo filo, che aprisse un poro nel ventricolo (F), che causasse la fuoriuscita di uno «spirito animale» attraverso un tubicino cavo, che gonfiasse i muscoli della gamba, che provocasse il ritrarsi del piede (Cartesio, 1664). Ma si trattava altrimenti di una buona idea. Lo stesso non si può dire della sua visione del ruolo della ghiandola pineale come

porta d'accesso alla coscienza (potremmo chiamarla la strozzatura cartesiana). Tale idea, il dualismo cartesiano, è irrimediabilmente sbagliata, come abbiamo visto nel capitolo 2. Ma benché il materialismo in una versione o nell'altra sia ora l'opinione corrente accettata quasi all'unanimità, non è difficile scoprire che anche i più sofisticati materialisti oggi si dimenticano spesso che una volta abbandonata la fantasmatica *res cogitans* di Cartesio, non esiste più un ruolo per una porta d'accesso centralizzata, o meglio per alcun centro *funzionale* del cervello. Non solo la ghiandola pineale non è la macchina con cui si inviano i fax all'anima e non è l'ufficio presidenziale del cervello, ma nemmeno lo sono altre porzioni specifiche del cervello stesso. Il cervello è il Quartier Generale, il luogo dove risiede l'osservatore ultimo, ma non c'è ragione di supporre che il cervello stesso abbia al suo interno un quartier generale più profondo, un santuario interno, al quale è necessario o sufficiente arrivare per avere un'esperienza cosciente. In breve, non esiste un osservatore all'interno del cervello.[2]

La luce viaggia molto più velocemente del suono, come ci ricorda l'esempio dei fuochi d'artificio, ma noi oggi sappiamo che il cervello impiega più tempo a elaborare gli stimoli visivi di quanto ne impieghi a elaborare quelli auditivi. Come il neuroscienziato Ernst Pöppel (1985, 1988) ha indicato, grazie a queste differenze che si controbilanciano, «l'orizzonte della simultaneità» è a *circa* dieci metri: la luce e il suono che lasciano lo stesso punto a circa dieci metri dagli organi di senso dell'osservatore producono delle reazioni neurali che sono «centralmente disponibili» nello stesso momento. Possiamo rendere più precisa questa misura? C'è un problema. Perché non basta misurare le distanze tra l'evento esterno e gli organi sensoriali, o le velocità di propagazione nei vari mezzi, o tener conto delle differenze individuali; bisogna anche – e ciò è fondamentale – decidere cosa considerare come «linea d'arrivo» nel cervello. Pöppel ha ottenuto il suo risultato comparando delle misure comportamentali: i tempi medi di reazione (la pressione di un pulsante) a degli stimoli auditivi e visivi. La differenza varia tra i 30 msec e i 40 msec, il tempo che impiega il suono a percorrere circa dieci metri (il tempo che impiega la luce a percorrere dieci metri è insignificante). Pöppel ha usato una linea d'arrivo periferica – il comportamento esterno – ma noi siamo naturalmente portati a pensare che l'*esperienza* della luce o del suono avvenga *tra* il momento in cui le vibrazioni colpiscono i nostri organi di senso e il momento in cui riusciamo a pigiare il

pulsante segnalando quell'esperienza. Essa si verifica, inoltre, in qualche luogo *centrale*, in qualche luogo nel cervello sul percorso eccitato che va dagli organi di senso al dito. Ci sembra che se potessimo dire *esattamente* dove avviene l'esperienza, potremmo dire esattamente quando avviene; e viceversa, se potessimo dire esattamente quando avviene, potremmo dire in quale luogo del cervello quell'esperienza cosciente si realizza.

L'idea che esista un tale luogo centrale nel cervello potrebbe essere chiamata *materialismo cartesiano*, poiché è la posizione a cui si arriva quando si scarta il dualismo cartesiano senza però abbandonare l'immagine di un Teatro centrale (ma materiale) dove «tutto converge». La ghiandola pineale potrebbe essere un candidato naturale per questo Teatro Cartesiano, ma ne sono stati proposti anche altri: il cingolato anteriore, la formazione reticolare, vari luoghi nei lobi frontali. Il materialismo cartesiano consiste nel ritenere che ci sia un confine o una linea d'arrivo cruciale in qualche parte del cervello e che l'ordine d'arrivo in quel sito equivale all'ordine con cui le esperienze «ci si presentano», perché *ciò che si verifica lì* è ciò di cui siamo coscienti. Probabilmente oggi nessuno sostiene esplicitamente il materialismo cartesiano e molti pensatori insisteranno nel dire che hanno esplicitamente rifiutato un'idea così ovviamente errata. Ma, come vedremo, le immagini persuasive connesse al Teatro Cartesiano ritornano a perseguitarci – profani e scienziati, in ugual misura – anche dopo che il suo fantasmatico dualismo sia stato smascherato ed esorcizzato.

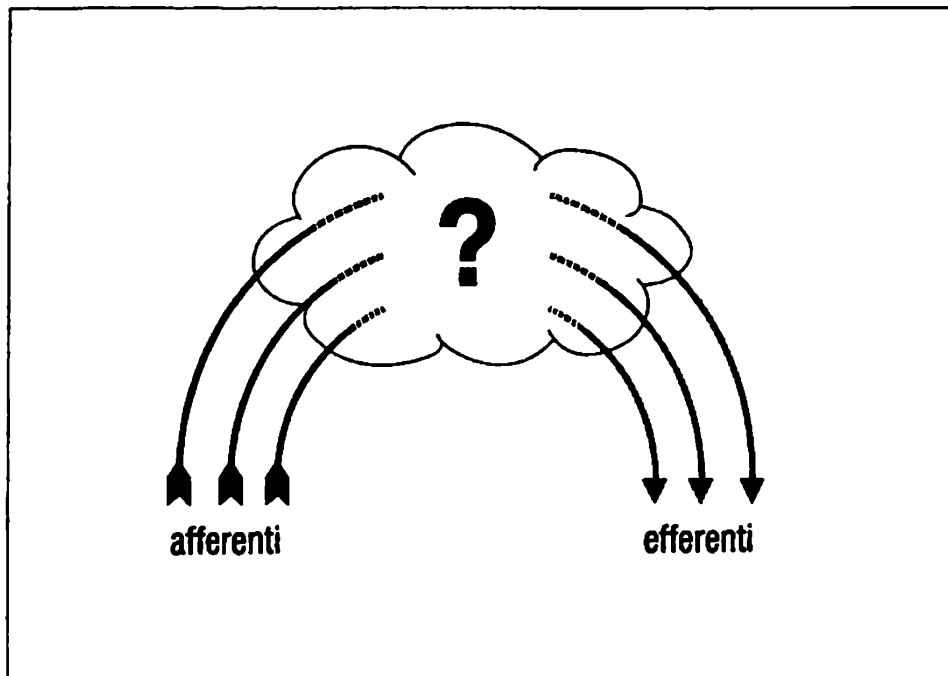
Il Teatro Cartesiano costituisce un'immagine metaforica del modo in cui l'esperienza cosciente debba risiedere nel cervello. Sembra essere dapprincipio un'innocente estrapolazione dal fatto familiare e innegabile che, *per gli intervalli di tempo macroscopici e usuali*, noi possiamo effettivamente classificare gli eventi nelle due categorie del «non ancora osservato» e del «già osservato». Ciò viene effettuato posizionando l'osservatore in un punto e rilevando i movimenti dei veicoli dell'informazione rispetto ad esso. Ma quando tentiamo di estendere questo metodo alla spiegazione di fenomeni che implicano intervalli temporali molto brevi, incontriamo una difficoltà *logica*: se il «punto» di vista dell'osservatore deve essere sparpagliato in una porzione piuttosto ampia del suo cervello, il suo senso soggettivo della sequenza e della simultaneità *deve* essere determinato da qualcosa che non sia «l'ordine d'arrivo», poiché l'ordine d'arrivo non è completamente definito fino a che non viene specificata la destinazione appropriata. Se A batte B su

un traguardo ma B batte A su un altro, quale risultato determina la sequenza soggettiva nella coscienza? (Cfr. Minsky, 1985, pp. 110-111). Pöppel parla del momento in cui la luce e il suono diventano «centralmente disponibili» nel cervello, ma quale punto (o punti) di «disponibilità centrale» dovrebbe essere considerato come quello (o quelli) che determina *l'ordine esperito*, e perché? Quando tentiamo di rispondere a questa domanda, siamo forzati ad abbandonare il Teatro Cartesiano e a sostituirlo con un nuovo modello.

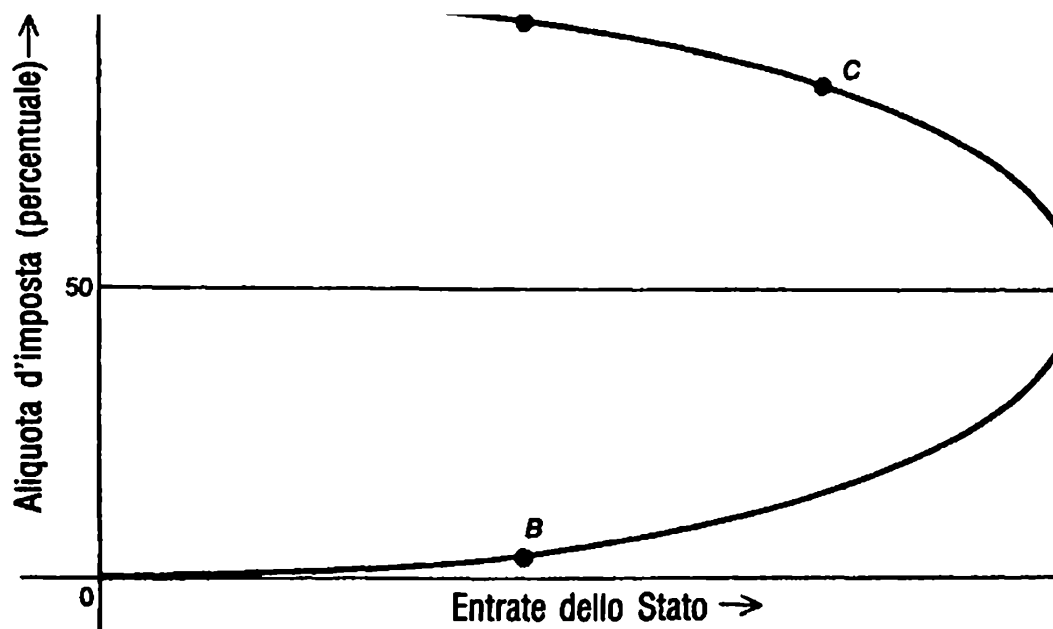
L'idea che esista un centro speciale nel cervello è una delle idee più tenaci e più ingannevoli tra quelle che frustrano i nostri tentativi di riflettere coerentemente sulla coscienza. Come vedremo, continua a riaffermarsi sotto nuove spoglie, per varie ragioni apparentemente irresistibili. Tanto per cominciare, c'è il nostro apprezzamento personale e introspettivo della «unità della coscienza», che ci fa tracciare la distinzione tra il «qui dentro» e il «lì fuori». Il confine tra «me» e «il mondo esterno» viene ingenuamente identificato con la mia pelle (e il cristallino dei miei occhi) ma, non appena ci accorgiamo che alcuni eventi nel nostro corpo possono essere inaccessibili «a noi», il grande esterno guadagna terreno. «Qui dentro» posso provare ad alzare il braccio, ma «lì fuori» non si muove perché «si è addormentato» o è paralizzato; le linee di comunicazione tra me (ovunque *io* sia) e il marchingegno neurale che controlla il mio braccio sono state manomesse. E se il mio nervo ottico fosse stato in qualche modo reciso non mi aspetterei di vedere, anche se i miei occhi fossero ancora intatti; avere delle esperienze visive è qualcosa che avviene evidentemente *al di qua* dei miei occhi, in qualche luogo tra i miei occhi e la mia voce quando ti dico cosa vedo.

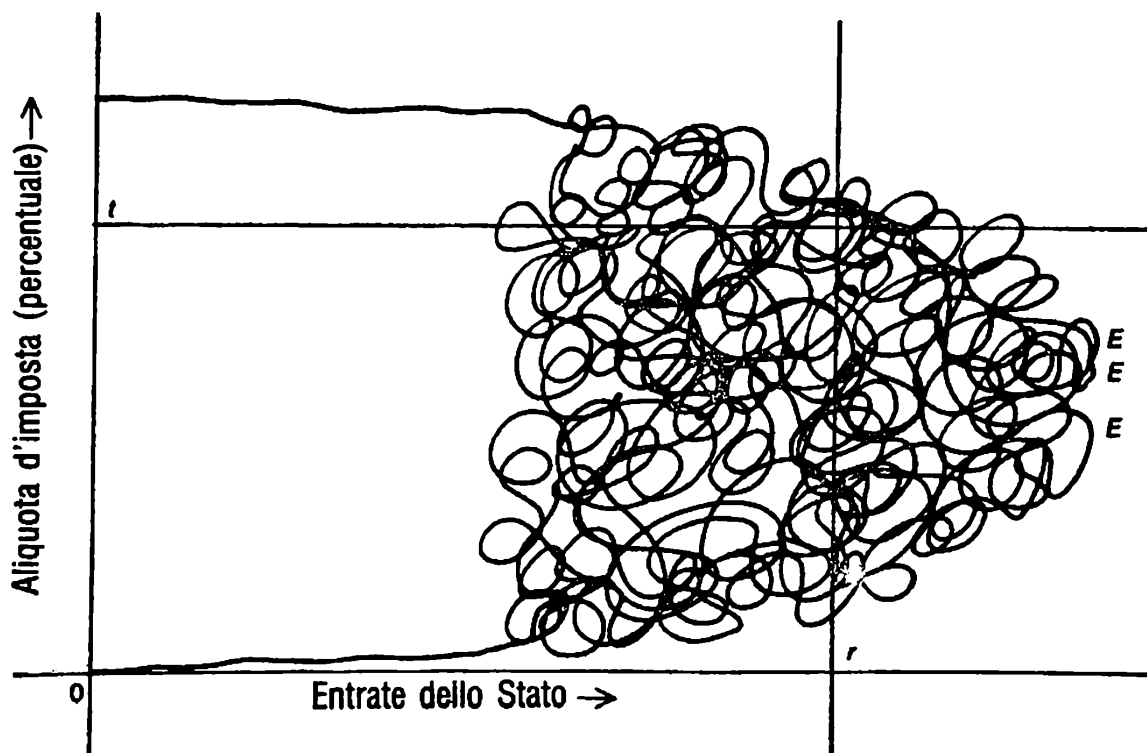
Non ne consegue *come una necessità geometrica* che la nostra mente cosciente sia collocata al *termine* di tutti i processi che si dirigono verso *l'interno*, immediatamente prima dell'*inizio* di tutti i processi che si dirigono verso *l'esterno* e realizzano le nostre azioni? Procedendo dalla periferia lungo i canali d'ingresso dall'occhio, per esempio, risaliamo il nervo ottico, poi ancora più su attraverso varie aree della corteccia visiva, e poi...? Procedendo dall'altro estremo periferico, navigando controcorrente, dai muscoli e dai neuroni motori che li controllano arriviamo all'area motoria supplementare nella corteccia e poi...? Questi due tragitti tendono uno verso l'altro risalendo due diversi versanti, l'afferente (l'ingresso) e l'efferente (l'uscita). Per quanto possa essere difficile determinare in pra-

tica dove si trovi esattamente lo «spartiacque cerebrale», non ci dovrebbe essere, ragionando in base ad una pura estrapolazione geometrica, un punto più alto, un punto di svolta, un punto siffatto che tutti gli interventi eseguiti su un lato di esso siano *precedenti all'esperienza*, e tutti gli interventi eseguiti sull'altro siano *successivi all'esperienza*?

Nell'immagine di Cartesio, ciò è visivamente ovvio poiché ogni cosa converge sulla stazione pineale o sgorga da essa. Sembrerebbe, allora, che, se fossimo impegnati nel realizzare un modello più attuale del cervello, dovremmo essere in grado di codificare le nostre esplorazioni con dei colori, usando, per esempio, il rosso per l'afferente e il verde per l'efferente: il punto mediano funzionale, il grande spartiacque mentale si troverebbe ovunque i nostri canali improvvisamente cambiano colore.

5.4

Questo argomento apparentemente stringente dovrebbe metterci in allarme. È il gemello di un altro argomento altrettanto falso che è stato sicuramente troppo influente negli ultimi anni: la nota curva di Arthur Laffer, che ha costituito il fondamento intellettuale (se così si può dire) della Reaganomics.

Se il governo fissa l'aliquota d'imposta allo 0%, il gettito fiscale sarà nullo; se la fissa al 100%, nessuno vorrà più lavorare per uno stipendio, ed anche così non otterrà alcun gettito; con il 2% otterrà approssimativamente il doppio che con l'1%, e così via; ma se l'aliquota cresce troppo il gettito fiscale comincerà a diminuire: le tasse si fanno onerose. Guardando all'altro

5.5 La curva di Laffer

estremo della scala, anche un'aliquota del 99% (come del 100%) può essere considerata più una confisca che una tassa, così difficilmente ci saranno più entrate; al 90% le cose vanno meglio per il governo, e ancor meglio andranno alla percentuale più invitante dell'80%. I versanti della curva considerata forse non sono così simmetrici, ma non dovrebbe esserci, per necessità geometrica, un punto in cui la curva gira, un'aliquota d'imposta che massimizza le entrate? Laffer era convinto che una diminuzione delle aliquote avrebbe in realtà prodotto un aumento del gettito fiscale, poiché le aliquote d'imposta si trovavano nel versante superiore della curva. Si trattava di un'idea attraente; a molti sembrò buona. Ma come Martin Gardner ha evidenziato, dal fatto che gli estremi della curva siano chiari non segue necessariamente che anche la parte sconosciuta e centrale della curva si debba comportare in modo regolare; e in modo satirico ha proposto in alternativa una «curva neo-lafferiana» che ha più di un «massimo», le cui comprensioni individuali dipendono da complessità della storia e delle circostanze che non possono essere spiegate da nessuna variazione di una singola variabile (Gardner, 1981). Noi dovremmo trarre la stessa morale su ciò che accade nella nebbia che si estende tra la periferia afferente e quella efferente: la chiarezza delle periferie non ci garantisce che la stessa distinzione continui ad esserci *in tutto il tragitto interno.* «Il groviglio tecnico» che Gardner immagina per l'economia rappresenta

5.6 La curva neo-lafferiana

la semplicità stessa se paragonata al guazzabuglio di attività che avvengono nelle regioni più centrali del cervello. Dobbiamo smettere di pensare che il cervello abbia un unico vertice funzionale o un punto centrale del genere. Tale supposizione non è un'innocua scorciatoia: è una cattiva abitudine. Per scrollarcela di dosso dobbiamo esaminare alcuni casi in cui questa cattiva abitudine di pensiero è in azione, ma abbiamo bisogno anche di una buona immagine con cui sostituirla.

2. *Il modello delle Molteplici Versioni*

Ecco una prima presentazione del sostituto, il modello delle Molteplici Versioni (*Multiple Drafts model*) della coscienza. Non mi aspetto certo che appaia immediatamente familiare e facile da visualizzare: l'idea del Teatro Cartesiano è così profondamente radicata in noi da farcelo apparire strano e contorto. Secondo il modello delle Molteplici Versioni, ogni tipo di percezione – in verità, ogni tipo di pensiero o attività mentale – è compiuto nel cervello da un processo parallelo e a piste multiple di interpretazione ed elaborazione dei dati sensoriali in ingresso. Le informazioni che entrano nel sistema nervoso sono sotto continua «revisione editoriale». Per esempio, poiché la nostra testa si muove un po' e i nostri occhi si muovono molto,

le immagini sulle nostre retine ondeggiano costantemente, come quelle dei filmetti amatoriali girati da persone che non riescono a tenere ferma la cinepresa. Ma a noi non sembra così. La gente spesso si stupisce nell'apprendere che i nostri occhi in condizioni normali saltellano da una parte all'altra in rapidi movimenti saccadici, fissando ogni punto per circa un quinto di secondo, e che tali movimenti, come quello della nostra testa, sono revisionati molto presto nel processo che va dal bulbo oculare alla... coscienza. Gli psicologi hanno imparato molto sui meccanismi tramite i quali raggiungiamo questi effetti normali ed hanno anche scoperto alcuni effetti speciali, come l'interpretazione della profondità in stereogrammi a punti casuali (Julesz, 1971).

5.7

Se osservi questi due quadrati leggermente differenti con uno stereoscopio (o semplicemente se li fissi cercando di incrociare leggermente lo sguardo fino a ottenere la fusione delle due immagini – alcune persone possono farlo senza l'aiuto di un congegno ottico), vedrai emergere una forma in tre dimensioni, grazie all'impressionante processo editoriale nel cervello che confronta e collaziona le informazioni provenienti da ciascun occhio. È possibile trovare la collimazione globalmente ottimale senza aver prima sottoposto ogni sequenza ordinata di dati a un elaborato processo di estrapolazione delle caratteristiche; ci sono sufficienti coincidenze di basso livello dei tratti salienti – i singoli puntini in uno stereogramma a punti casuali – per determinare una soluzione.

I processi editoriali del cervello hanno bisogno di un tempo piuttosto lungo per produrre questi effetti speciali; in altri casi, invece, le cose sono molto più rapide, come nel caso dell'effetto

McGurk (McGurk e Macdonald, 1979). Quando un film straniero è doppiato nella propria lingua, il più delle volte gli spettatori non si accorgono delle discrepanze tra il movimento delle labbra che vedono e i suoni che ascoltano – a meno che il doppiaggio non sia fatto sciattamente. Ma che succede nel caso di un sonoro che si accordi perfettamente con le immagini tranne che nel caso voluto di alcune consonanti fuori posto? (Per riprendere un vecchio esempio, supponiamo che le labbra della persona filmata dicano «da sinistra a destra» e che la voce nel sonoro dica «da dinistra a sestra».) Di cosa avrà esperienza la gente? *Ascolterà* «da sinistra a destra». Nella contesa editoriale, artificialmente indotta, tra i contributi degli occhi e quelli degli orecchi, gli occhi vincono – in questo caso.[3]

Questi processi editoriali avvengono in ampie frazioni di secondo; in questo lasso di tempo possono verificarsi varie aggiunte, fusioni, correzioni e riscritture del contenuto, in vario ordine. Noi non abbiamo una diretta esperienza di quello che avviene sulle nostre retine, nei nostri orecchi, sulla superficie della nostra pelle. Nella nostra effettiva esperienza rientra solo il prodotto finito di questi svariati processi di interpretazione – che sono, in effetti, processi editoriali. Essi si servono di rappresentazioni relativamente grezze e unilaterali per produrre delle rappresentazioni collazionate, revisionate, migliorate che si realizzano nel flusso dell'attività che si svolge in varie parti del cervello. Tutto ciò è riconosciuto praticamente da ogni teoria della percezione, ma ecco la novità del modello delle Molteplici Versioni: la rilevazione o discriminazione delle caratteristiche *deve essere fatta solo una volta*. Cioè, una volta che una particolare «osservazione» di qualche caratteristica è stata effettuata da una porzione specializzata e delimitata del cervello, non è necessario che il contenuto dell'informazione così definito sia inviato altrove per essere discriminato da qualche «discriminatore capo». In altre parole, la discriminazione non conduce a una ri-*presentazione* della caratteristica già discriminata per il piacere del pubblico nel Teatro Cartesiano – poiché non c'è alcun Teatro Cartesiano.

Queste definizioni di contenuto distribuite spazialmente e temporalmente nel cervello sono precisamente localizzabili sia nel tempo che nello spazio, ma il loro insorgere *non* segna l'insorgere della coscienza del loro contenuto. Non è possibile stabilire se un particolare contenuto così discriminato comparirà alla fine come un elemento nell'esperienza cosciente; e genera confusione, come vedremo, chiedersi *quando esso sia dive-*

nuto cosciente. Queste discriminazioni distribuite di contenuto conducono, nel corso del tempo, a qualcosa che è *abbastanza simile* ad un flusso o a una sequenza narrativa, che, si può pensare, è soggetta ad una continua revisione da parte di molti processi distribuiti nel cervello che continuano a operare incessantemente e indefinitamente nel futuro. Questo flusso di contenuti assomiglia solo in parte ad una narrazione a causa della sua molteplicità: in ogni momento ci sono molteplici «versioni» di frammenti narrativi a vari livelli di revisione in vari luoghi del cervello.

Sondare questo flusso in luoghi e tempi differenti produce effetti differenti, precipita narrazioni differenti da parte del soggetto. Se si procrastina troppo il sondaggio (si fa passare una notte, per esempio), si rischia molto probabilmente di non trovare più una narrazione – oppure di trovarne una così rimasticata o «razionalmente ricostruita» da aver perso la sua integrità. Se si sonda «troppo presto», si possono raccogliere utili dati sulla velocità con la quale una particolare discriminazione viene compiuta dal cervello, ma al prezzo di distorcere il corso normale del flusso molteplice. La cosa più importante, comunque, è che il modello delle Molteplici Versioni evita l'allettante errore di supporre che ci debba essere una narrazione unica e canonica (la versione «finale» o «pubblicata», potremmo dire) – che costituisce il *reale* flusso di coscienza del soggetto, indipendentemente dal fatto che lo sperimentatore (o perfino il soggetto stesso) possa o non possa accedere ad esso.

Finora ti sembrerà probabilmente che questo modello non abbia nulla a che fare con la coscienza come la conosci per esperienza personale. Ma ciò è dovuto al fatto che sei ancora abituato a pensare alla coscienza come a uno spettacolo che si svolge nel Teatro Cartesiano. Per scrollarti di dosso questa comoda e naturale abitudine e trasformare il modello delle Molteplici Versioni in un'alternativa vivida e credibile, ci vorrà molto impegno, e un impegno piuttosto bizzarro. Questa sarà sicuramente la parte più difficile del libro, ma è essenziale all'economia generale della teoria e non può essere saltata! Non c'è matematica, grazie a Dio. Devi solo pensare attentamente e vividamente, impedendo che nella tua mente si formi un'immagine seducente ed errata al posto di quella giusta. Incontrerai una moltitudine di semplici esperimenti mentali che aiuteranno la tua immaginazione nel corso di questo insidioso cammino. Quindi preparati a qualche faticoso esercizio; alla fine ti si svelerà una nuova prospettiva sulla coscienza, che implica una profonda riforma (ma non una radicale rivoluzione) del nostro

modo di pensare al cervello. (Per un modello simile, vedi il modello della coscienza come «tessitura di scenari» di William Calvin [1989].)

Un buon modo pèr riuscire a comprendere una teoria è quello di vedere come affronta un fenomeno relativamente semplice che non riesce ad essere spiegato dalla vecchia teoria. La prima prova riguarda una scoperta sul movimento apparente che è stata provocata, sono felice di dirlo, dalle domande di un filosofo. La cinematografia e la televisione si basano sulla creazione di un movimento apparente ottenuto presentando in rapida successione delle fotografie «immobili»; sin dai primordi dell'era del cinema, gli psicologi hanno studiato questo fenomeno, chiamato movimento *phi* da Max Wertheimer (1912), il primo a studiarlo sistematicamente. Nel caso più semplice, se due o più punti separati da non più di 4 gradi nell'angolo visivo vengono brevemente accesi in rapida successione, a noi sembrerà di vedere un unico punto che si muove avanti e indietro. Il movimento *phi* è stato studiato in vari casi e uno dei più sorprendenti viene riferito dagli psicologi Paul Kolers e Michael von Grünau (1976). Il filosofo Nelson Goodman aveva chiesto a Kolers se il fenomeno *phi* continuava a sussistere anche nel caso di due punti illuminati di colore differente e, se sì, cosa accadeva al colore «del punto» durante il «movimento». L'illusione del movimento sarebbe svanita e al suo posto ci sarebbero stati due punti lampeggianti separati? Il punto che illusoriamente «si muove» avrebbe cambiato gradualmente colore tracciando una traiettoria attraverso il solido dei colori (la sfera tridimensionale che rappresenta tutte le tonalità cromatiche)? (Forse vuoi fare la tua previsione prima di continuare a leggere.) La risposta, quando Kolers e von Grünau eseguirono gli esperimenti, fu inaspettata: due punti luminosi differentemente colorati furono accesi per 150 msec ciascuno (con un intervallo di 50 msec); il primo punto sembrava cominciare a muoversi e poi cambiare colore repentinamente *a metà della traiettoria illusoria* verso il secondo luogo. Goodman si chiese: «Come è possibile che riusciamo [...] ad inserire il punto luminoso negli spazi e nei tempi che intercorrono nel tragitto che va dal primo al secondo punto *prima che il secondo punto luminoso si accenda*?». (Goodman, 1978, p. 73.)

La stessa domanda, naturalmente, può essere sollevata rispetto a qualunque fenomeno *phi*, ma quello colorato di Kolers solleva vividamente il problema. Supponiamo che il primo punto luminoso sia rosso e il secondo verde. A meno che non si verifichi una «precognizione» nel cervello (un'ipotesi stravagante che cer-

cheremo di rimandare indefinitamente), il contenuto illusorio, «*il rosso che diventa verde a metà del percorso*», non può essere creato *fino a che* una qualche identificazione del secondo punto, verde, non si verifichi nel cervello. Ma se il secondo punto luminoso è già «nell'esperienza cosciente», non è troppo tardi per inserire tale contenuto illusorio tra l'esperienza cosciente del punto rosso e l'esperienza cosciente del punto verde? Come riesce il cervello ad eseguire questo gioco di prestigio?

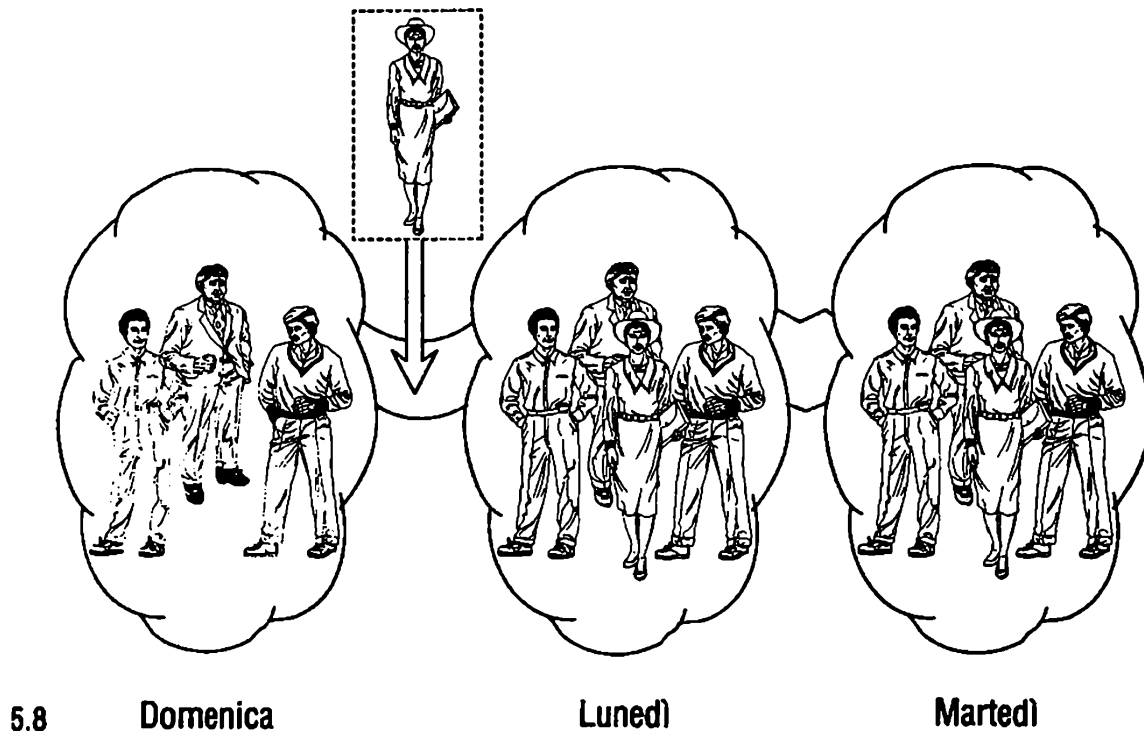
Il principio secondo cui le cause devono precedere gli effetti si applica ai processi multipli e distribuiti che realizzano il lavoro editoriale nel cervello. Ogni processo specifico che necessita informazioni da qualche fonte deve infatti attendere per avere quelle informazioni: fino a che non le ottiene non può andare avanti. Per questo ogni spiegazione «magica» o precognitiva del cambiamento di colore nel fenomeno *phi* deve essere esclusa. Il contenuto *punto verde* non può essere attribuito a nessun evento, conscio o inconscio, fino a che la luce che parte dal punto verde non abbia raggiunto l'occhio e messo in movimento la normale attività neurale nel sistema visivo fino al livello in cui la discriminazione del verde viene eseguita. Così, la discriminazione (illusoria) del «rosso che diventa verde» deve essere eseguita *dopo* la discriminazione del punto verde. Ma, allora, poiché ciò di cui hai esperienza cosciente è *prima rosso, poi rosso che diventa verde e quindi verde*, ne consegue («certamente») che la coscienza dell'intero evento deve essere rimandata fino a che il punto verde non sia stato (inconsciamente?) percepito. Se trovi questa conclusione stringente, sei ancora imprigionato nel Teatro Cartesiano. Un esperimento mentale ti aiuterà ad uscirne.

3. *Revisioni orwelliane e staliniane*

> Non sono davvero sicuro se gli altri non riescano a percepirmi o se, una frazione di secondo dopo che la mia faccia interferisce con il loro orizzonte, un milionesimo di secondo dopo che hanno gettato il loro sguardo su di me, io inizi già ad essere spazzato via dalla loro memoria: dimenticato prima di arrivare all'avaro e triste arcangelo della rimembranza.
>
> ARIEL DORFMAN, *Mascara*, 1988

Supponiamo che io riesca a manomettere il tuo cervello e a inserire nella tua memoria una fantomatica donna con cappello

lì dove non c'era (per esempio, alla festa di domenica). Anche se lunedì, quando ti ritorna in mente la festa, ti ricordi di lei e non riesci a trovare nessuna risorsa interna che ti permetta anche solo di dubitare della veracità della tua memoria, noi dovremmo continuare a dire tuttavia che *non* hai mai avuto un'esperienza di lei; sicuramente non alla festa di domenica. Naturalmente la tua successiva esperienza di un ricordo (falso) può essere vivida quanto vuoi e martedì potremmo essere tutti d'accordo che hai avuto una vivida esperienza cosciente di una donna con cappello alla festa, ma la *prima* esperienza, dovremmo insistere, l'hai avuta lunedì e non domenica (benché a te non sembri così).

5.8 Domenica Lunedì Martedì

Noi non siamo in grado di inserire finte memorie tramite neurochirurgia, ma talvolta le nostre memorie si prendono gioco di noi, e così ciò che non può essere realizzato chirurgicamente accade spontaneamente nel cervello. Talvolta ci sembra di ricordare, anche vividamente, delle esperienze che non abbiamo mai avuto. Tali contaminazioni post-esperienziali o revisioni di memorie potrebbero essere chiamate orwelliane, perché ricordano l'agghiacciante descrizione che George Orwell dà nel suo romanzo *1984* del Ministero della Verità, che febbrilmente riscriveva la storia negando così alla posterità ogni accesso al (reale) passato.

La possibilità di una revisione (orwelliana) post-esperien-

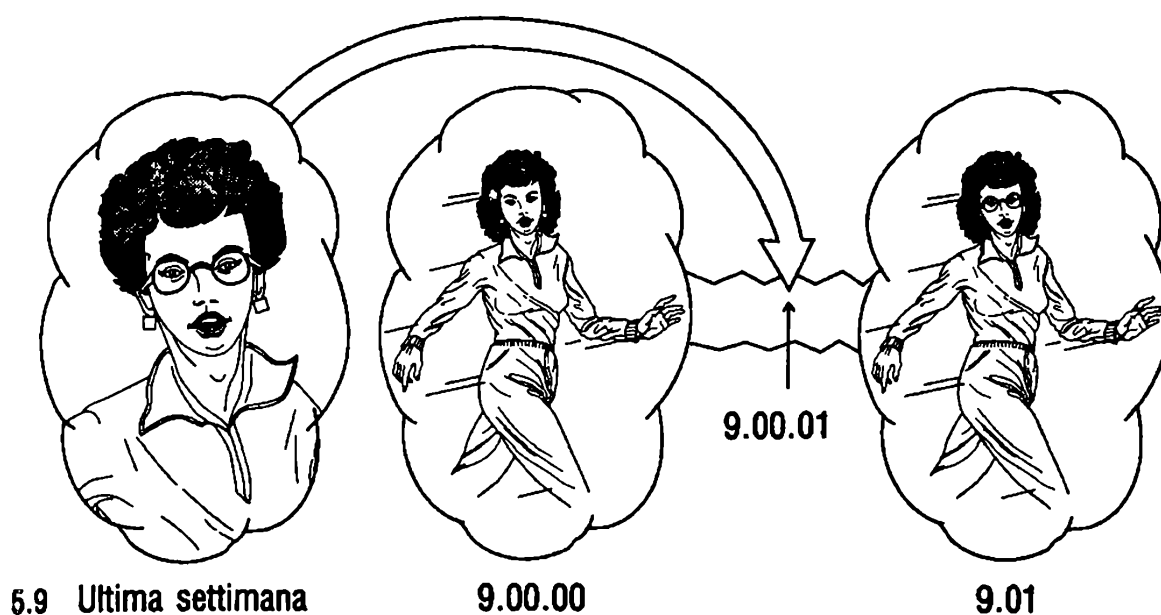
ziale mette in luce una delle nostre più fondamentali distinzioni: quella tra apparenza e realtà. Poiché riconosciamo la possibilità (almeno teorica) di una revisione orwelliana, riconosciamo il rischio insito nell'inferire «questo è quello che è davvero accaduto» da «questo è quello che ricordo» e quindi resistiamo – con buone ragioni – ad ogni diabolico «operazionalismo» che cerca di convincerci che ciò che ricordiamo (o ciò che gli storiografi registrano nei loro archivi) è *proprio* ciò che è davvero accaduto.[4]

Una revisione orwelliana non è l'unico modo per ingannare i posteri. Un altro consiste nell'inscenare processi farseschi, presentando trascrizioni accurate di false testimonianze e finte confessioni, e integrandole con prove abilmente contraffatte. Un tale stratagemma si potrebbe chiamare *staliniano*. Si noti che se di solito siamo sicuri di quale modo di falsificazione sia stato tentato ai nostri danni, quello orwelliano o quello staliniano, ciò dipende da una regia approssimativa; in ogni campagna di disinformazione *riuscita*, si è incapaci di dire se i resoconti giornalistici siano delle descrizioni orwelliane di processi mai avvenuti o vere descrizioni di processi farseschi realmente avvenuti. Se *tutte* le tracce – giornali, videoregistrazioni, epitaffi, memorie personali, testimoni viventi – sono state occultate o alterate, non avremmo strumenti per sapere se la fabbricazione è avvenuta *prima* e si è conclusa con un processo farsesco di cui abbiamo un resoconto accurato, o se piuttosto, *dopo* una sommaria esecuzione, c'è stata una fabbricazione storiografica che ha cancellato il misfatto, ma nessun tipo di processo si è mai svolto *davvero*.

La distinzione tra metodi orwelliani e metodi staliniani per la produzione di archivi fuorvianti funziona senza problemi nel mondo di tutti i giorni, in dimensioni temporali macroscopiche; e si potrebbe pensare che si applichi tranquillamente *in ogni situazione*, ma questa è un'illusione, come vedremo tra un istante in un esperimento ideale che si differenzia dal precedente solo per le dimensioni temporali.

Supponiamo che te ne stai in disparte e una donna con i capelli lunghi ti passa velocemente di fronte. Circa un secondo *dopo*, una memoria sotterranea di qualche donna precedente – una donna con i capelli corti e gli occhiali – contamina la memoria di ciò che hai appena visto; quando un minuto più tardi ti chiedono di descrivere la donna che hai appena visto, tu affermi, sinceramente ma erroneamente, che aveva gli occhiali. Proprio come nel caso della donna con il cappello alla festa, noi

siamo portati a dire che la tua *originale esperienza visiva*, ben distinta dalla memoria che la segue di qualche secondo, *non* riguardava una donna con gli occhiali. Ma a causa di una successiva contaminazione della memoria, ti sembra senz'altro che, fin dal primo momento che l'hai vista, ti ha colpito il

5.9 Ultima settimana 9.00.00 9.01

fatto che avesse gli occhiali. Si è verificata una revisione orwelliana: c'è stato un fugace momento, prima della contaminazione della memoria, in cui *non* ti sembrava che avesse gli occhiali; durante quel breve istante, il contenuto *reale* della tua esperienza cosciente è stato una donna con i capelli lunghi e *senza* occhiali, ma questo fatto storico è diventato inerte, non ha lasciato tracce, grazie alla contaminazione della memoria che è avvenuta un secondo dopo che l'hai intravista.

Questa spiegazione dell'accaduto può comunque essere smentita da una descrizione differente. Le tue anteriori memorie sotterranee di quella donna con gli occhiali potrebbero benissimo aver contaminato la tua esperienza *nella fase afferente*, nel corso dell'elaborazione dell'informazione che avviene «anteriormente alla coscienza», cosicché si può dire che fin dall'inizio la tua esperienza ha avuto un contenuto, gli occhiali, frutto di una sorta di *allucinazione*. In questo caso, la tua memoria ossessiva di una precedente donna con gli occhiali giocherebbe un tiro staliniano ai tuoi danni, inscenando una processo farsesco nella tua esperienza, che tu poi ricordi fedelmente, grazie agli archivi della tua memoria. A prima vista questi due casi sembrerebbero perfettamente distinti. Nella prima descrizione (figura 5.9), non hai avuto alcuna allucinazione quando la donna ti è passata di

5.10 Ultima settimana 9.00.00 9.01

fronte, ma hai sofferto di un'allucinazione mnemonica successivamente: hai false memorie della tua effettiva («reale») esperienza. Nella seconda (figura 5.10), hai avuto un'allucinazione quando la donna ti è passata di fronte, e successivamente ti ricordi fedelmente di quell'allucinazione (che «è realmente avvenuta nella coscienza»). Queste descrizioni non rappresentano forse delle possibilità perfettamente distinte anche in frazioni di tempo piccolissime?

No. Quando ci occupiamo di dimensioni temporali microscopiche la distinzione tra revisione percettiva e revisione della memoria, che funziona nettamente su scala più grande, non è più garantita e non è detto che abbia ancora senso. Siamo entrati nella zona nebulosa in cui il punto di vista del soggetto è sfocato sia spazialmente che temporalmente e in cui la domanda *«orwelliano o staliniano?»* perde la sua forza.

C'è una finestra temporale che si apre quando la donna con i capelli lunghi appare fuggevolmente, eccitando la tua retina, e si chiude quando tu esprimi – a te stesso o a qualcun altro – la tua convinzione finale che aveva gli occhiali. In qualche momento durante quest'intervallo, il contenuto *con occhiali* si è aggiunto in modo spurio al contenuto *donna dai capelli lunghi.* Potremmo presumere (e potremmo un giorno perfino confermarlo in dettaglio) che ci sia stato un breve momento in cui solo il contenuto *donna dai capelli lunghi* sia stato discriminato nel cervello, e che il contenuto *con occhiali* sia stato erroneamente «aggiunto» ad esso *solo successivamente.* In effetti, sarebbe plausibile supporre che proprio tale discriminazione di una donna con i capelli lunghi sia stato ciò che ha fatto scattare il ricordo di una donna precedente con gli occhiali. Ciò che non potremmo

sapere, però, è se l'aggiunta spuria sia avvenuta «prima o dopo il fatto» – il fatto presunto della «reale esperienza cosciente». Sei stato prima cosciente di una donna con i capelli lunghi e senza occhiali e poi sei stato cosciente di una donna con i capelli lunghi e con gli occhiali (e questo secondo contenuto della coscienza ha spazzato via dalla memoria l'esperienza precedente) o fin dal primo istante la tua esperienza cosciente è stata contaminata dagli occhiali?

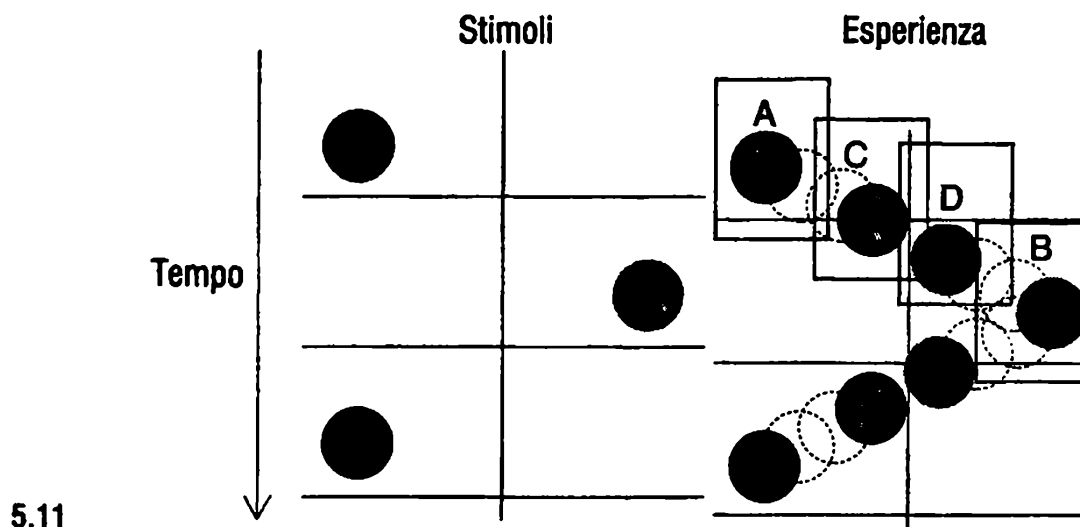
Se il materialismo cartesiano fosse vero, questa questione dovrebbe avere una risposta, anche se né noi – né tu – potremmo mai determinarla tramite alcun esame. Giacché il contenuto che «ha superato per primo la linea d'arrivo» è stato o *donna dai capelli lunghi* o *donna dai capelli lunghi con occhiali*. Ma quasi tutti i pensatori insisterebbero nel dire che il materialismo cartesiano è falso. Ciò che non hanno riconosciuto, però, è che questo implica che non esista alcuna linea d'arrivo privilegiata, cosicché l'ordine temporale delle discriminazioni non può essere ciò che fissa l'ordine soggettivo nell'esperienza. Questa conclusione non è facile da accettare, ma possiamo renderla più attraente esaminando le difficoltà in cui ci si va a cacciare se si rimane attaccati all'alternativa tradizionale.

Consideriamo il fenomeno *phi* colorato di Kolers. I soggetti *riferiscono* di aver visto un cambiamento di colore del punto mobile a metà della traiettoria dal rosso al verde. Questo breve testo eterofenomenologico è stato raffinato da Kolers grazie ad un ingegnoso espediente: i suoi soggetti usavano un congegno puntatore che dovevano «sovrapporre» *retrospettivamente ma il più presto possibile* alla traiettoria dell'illusorio punto mobile; compiendo così un atto linguistico dal contenuto «il punto luminoso ha cambiato il suo colore qui» (Kolers e von Grünau, 1976, p. 330).

Cosicché nel mondo eterofenomenologico dei soggetti c'è un cambiamento di colore a metà traiettoria, e l'informazione che indica a quale colore passare (e in quale direzione muoversi) deve provenire da qualche parte. Ricordati della formulazione di Goodman dell'enigma: «Come è possibile che riusciamo [...] a inserire il punto luminoso negli spazi e nei tempi che intercorrono nel tragitto che va dal primo al secondo punto *prima che il secondo punto luminoso si accenda*?». Forse, come hanno pensato alcuni psicologi, l'informazione proviene da *esperienze precedenti*; forse, come il cane di Pavlov che giunge ad aspettarsi il cibo ogni qual volta il campanello suona, così questi soggetti sono giunti ad aspettarsi di vedere il secondo punto ogni qual volta

vedono il primo punto, e in forza dell'abitudine si rappresentano davvero il passaggio prima di ricevere una qualche informazione dal caso particolare in atto. Ma quest'ipotesi è stata falsificata. Anche alla prima prova (cioè, senza alcuna possibilità di condizionamento precedente), la gente è soggetta al fenomeno *phi*. Inoltre, in prove successive, la direzione e il colore del secondo punto possono essere casualmente modificati senza far scomparire l'effetto. Così l'informazione proveniente dal secondo punto luminoso (colore e posizione) deve in qualche modo essere usata dal cervello per creare la versione «revisionata e pubblicata» dai soggetti.

Consideriamo, prima, l'ipotesi che ci sia un meccanismo staliniano. Nella sala di montaggio del cervello, collocata prima della coscienza, si verifica una dilazione, allo stesso modo in cui nel nastro usato per trasmettere i programmi «in diretta» c'è un anello sovrabbondante che dà ai censori nella stanza di controllo qualche secondo per coprire con un beep le oscenità prima di trasmettere il segnale. *Nella sala di montaggio* arriva per primo il fotogramma A, quello del punto rosso, e poi quando arriva il fotogramma B, quello del punto verde, vengono creati alcuni fotogrammi intermedi (C e D) che vengono incorporati nel film che sarà proiettato nel teatro della coscienza (nell'ordine A, C, D, B). Quando il «prodotto finito» arriva alla coscienza, presenta già la sua inserzione illusoria.

5.11

L'ipotesi alternativa è che ci sia un meccanismo orwelliano. Un attimo dopo la coscienza del primo *e* del secondo punto (senza alcuna illusione di movimento), una specie di addetto alla revisione storica, nell'ufficio del protocollo degli archivi della memoria del cervello, si rende conto che la storia nuda e cruda in questo caso non è molto credibile, e così reinterpreta

gli eventi base, il rosso seguito dal verde, e confeziona una narrazione sull'intera vicenda, che comprende il percorso intermedio e il cambiamento di colore a metà tragitto; e installa questa storia, che comprende le sue glosse – i fotogrammi C e D (figura 5.11) – nell'archivio della memoria per ogni futura necessità. Poiché tale addetto lavora molto velocemente, nello spazio di una frazione di secondo – il tempo necessario a confezionare (ma non a pronunciare) un resoconto verbale di ciò di cui tu hai avuto esperienza – il fascicolo su cui ti basi, custodito nell'archivio della memoria, è già contaminato. Tu *dici* e *credi* di aver visto un movimento e un cambiamento di colore illusori, ma in realtà questa è un'allucinazione della memoria, non un accurato ricordo di una tua originaria esperienza cosciente.

Come possiamo sapere quale di queste due ipotesi è corretta? Sembrerebbe possibile escludere l'ipotesi staliniana molto facilmente, a causa della postulata dilazione nella coscienza. Nell'esperimento di Kolers e von Grünau, c'era una differenza di 200 msec tra l'accensione del punto rosso e quella del punto verde e poiché, *ex hypothesi*, l'*intera esperienza* non può essere composta nella sala di montaggio prima che il contenuto *punto verde* sia arrivato nella sala stessa, la coscienza del punto rosso iniziale dovrà essere stata differita di almeno quel tempo. (Se la sala di montaggio invia il contenuto *punto rosso* al teatro della coscienza immediatamente, prima di ricevere il fotogramma B, e poter creare quindi i fotogrammi C e D, il soggetto presumibilmente avrà l'esperienza di una lacuna nel film, di una dilazione di almeno 200 msec tra A e C – altrettanto rilevabile quanto la mancanza di una sillaba in una parola o di cinque fotogrammi in un film.)

Supponiamo di aver chiesto ai soggetti di premere un pulsante «non appena hanno l'esperienza di un punto rosso». Troveremo una differenza minima o nulla nei tempi di risposta nel caso di un punto rosso da solo e in quello di un punto rosso seguito 200 msec più tardi da uno verde (in quest'ultimo caso i soggetti riferiscono il cambiamento di colore nel movimento apparente). Ciò non potrebbe essere dovuto al fatto che c'è *sempre* una dilazione di almeno 200 msec nella coscienza? No, ci sono prove sufficienti per affermare che le risposte sotto controllo cosciente, benché più lente delle risposte costituite da riflessi involontari, come un battito di palpebre, si verificano con le minime latenze (dilazioni) fisicamente possibili. Dopo aver sottratto il tempo calcolabile in cui gli impulsi nervosi

compiono il viaggio di andata e ritorno tra le terminazioni sensorie e il sistema nervoso centrale e il tempo di preparazione delle risposte, non rimane un tempo sufficientemente lungo nella «elaborazione centrale» da poterci infilare una dilazione di 200 msec. Così la risposta (premere il pulsante) deve essere stata iniziata prima della discriminazione del secondo stimolo, il punto luminoso verde.

Con ciò la vittoria dell'ipotesi orwelliana, di un meccanismo di revisione post-esperienziale, sembrerebbe assicurata: non appena il soggetto è divenuto cosciente del punto rosso, ha cominciato a premere il pulsante. *Mentre quel pulsante veniva premuto*, egli è diventato cosciente del punto verde. Poi entrambe queste esperienze sono state spazzate via dalla memoria e sostituite da un documento falsificato che parla di un punto rosso che si muove e diventa verde a metà strada. Il soggetto riferisce prontamente e sinceramente ma *falsamente* di aver visto un punto rosso che si muoveva verso il punto verde prima di cambiare colore. Se il soggetto insiste nel dire che era cosciente fin dall'inizio del movimento del punto rosso e del cambiamento di colore, lo psicologo orwelliano gli spiegherà con sicurezza che si sbaglia: la sua memoria lo sta ingannando, il fatto di aver premuto il pulsante in quell'esatto momento è una prova conclusiva del fatto che era cosciente di un punto rosso (immobile) prima che si presentasse il punto verde. Dopo tutto le istruzioni erano di premere il pulsante *quando fosse stato cosciente* del punto rosso. Ed egli deve essere stato cosciente del punto rosso circa 200 msec prima che possa essere stato cosciente del suo movimento e del cambiamento di colore. Se a lui non sembra così, è un suo errore.

Il difensore dell'alternativa staliniana, però, non si sente per questo sconfitto. In realtà, egli sostiene, il soggetto ha risposto al punto rosso *prima* di esserne cosciente! Le indicazioni date al soggetto (di rispondere al punto rosso) sono in qualche modo trapelate dalla coscienza nella sala di montaggio, che (inconsciamente) ha messo in moto la pressione del pulsante prima di aver inviato la versione montata (fotogrammi ACDB) alla coscienza per la «prima visione». La memoria del soggetto non lo ha affatto ingannato; egli sta riferendo esattamente ciò di cui è stato cosciente, a parte la sua insistenza nel dire che ha premuto coscientemente il pulsante dopo aver visto il punto rosso; la sua «prematura» pressione del pulsante è stata inconsciamente (o preconsciamente) scatenata.

Laddove la teoria staliniana postula la reazione di premere

un pulsante ad un rilevamento *in*conscio di un punto rosso, la teoria orwelliana postula un'esperienza *cosciente* di un punto rosso che viene immediatamente cancellata dalla memoria da ciò che segue. Così ecco lo scoglio: abbiamo due modelli differenti di ciò che accade nel fenomeno *phi* colorato. Uno postula un «inserimento» staliniano sul percorso ascendente, pre-esperienziale; l'altro una «revisione di memoria» orwelliana sul persorso discendente, post-esperienziale; ed *entrambi* sono coerenti con *qualunque* cosa il soggetto possa dire, pensare o ricordare. Nota che l'incapacità a distinguere queste due possibilità non riguarda solo l'*osservatore esterno* al quale, si potrebbe pensare, manca qualche dato privato di cui è invece in possesso il soggetto, grazie al suo «accesso privilegiato». Tu, come soggetto di un esperimento sul fenomeno *phi*, *non potresti* scoprire nulla nell'esperienza dalla tua prospettiva personale in prima persona che potrebbe essere a favore dell'una o dell'altra teoria; l'esperienza ti «sembrerebbe la stessa» in entrambe le spiegazioni.

È davvero così? E se provassi a concentrare tutta la tua attenzione sull'esperienza, non riusciresti a scorgere una qualche differenza? Supponiamo che lo sperimentatore ti renda le cose più facili, rallentando l'esposizione, allungando gradualmente l'intervallo tra lo stimolo rosso e quello verde. È ovvio che se l'intervallo si fa troppo lungo tu puoi riconoscere la differenza tra *percepire* il movimento e *inferire* il movimento. (È una notte buia e tempestosa; al primo lampo mi vedi sulla tua sinistra; due secondi dopo c'è un altro lampo e mi vedi alla tua destra. Tu inferisci che mi devo essere mosso, e sicuramente sei in grado di dire che in questa circostanza stai solo inferendo il movimento, non mi *hai visto* muovere.) Man mano che lo sperimentatore allunga l'intervallo tra gli stimoli, tu riesci sempre meglio a stabilire questa differenza. Cominci a dire cose come:

«Questa volta non mi sembrava che il punto rosso si muovesse, ma dopo aver visto il punto verde ho l'impressione che il punto rosso si sia mosso nella sua direzione e abbia cambiato colore».

In effetti, c'è una serie intermedia di intervalli in cui la fenomenologia è piuttosto paradossale: vedi i punti luminosi come due lampi immobili *e* come una cosa che si muove! Questa specie di movimento apparente è facilmente distinguibile da quello più veloce e più fluido che vediamo nei film e nella televisione, ma

la nostra capacità di riconoscere *questa* differenza è irrilevante nella disputa tra lo psicologo orwelliano e quello staliniano. Essi sono d'accordo nel dire che nelle appropriate condizioni tu puoi riconoscere questa differenza. Ciò su cui si trovano in disaccordo è il modo in cui descrivere i casi di movimento apparente che tu *non riesci* a distinguere dal movimento reale – i casi in cui tu *percepisci* realmente il movimento illusorio. Per dirla in modo spicciolo, in questi casi è la tua memoria che si prende gioco di te oppure sono i tuoi occhi?

Ma anche se tu, il soggetto, non puoi dire se questo fenomeno sia staliniano o orwelliano, non potrebbero gli scienziati – gli osservatori esterni – trovare qualcosa nel tuo cervello che ci illumini sul caso? Qualcuno potrebbe essere spinto ad escludere questa ipotesi come *inconcepibile*. «Prova solo ad immaginare qualcun *altro* che conosca meglio di te quello di cui tu sei cosciente! Impossibile!» Ma è davvero inconcepibile? Osserviamo più attentamente. Supponiamo che questi scienziati abbiano a disposizione delle informazioni davvero accurate (raccolte tramite varie tecnologie d'analisi cerebrale) sugli esatti «tempi d'arrivo» o «di creazione» di ogni rappresentazione, di ogni veicolo del contenuto, in ogni luogo nel tuo sistema nervoso. Ciò permetterebbe loro di conoscere il momento preciso in cui puoi *iniziare* a reagire in qualunque maniera – conscia o inconscia – a qualsiasi contenuto particolare (escludendo la precognizione miracolosa). Ma il momento *effettivo* in cui tu sei *diventato cosciente* di quel contenuto (se mai lo sei diventato) potrebbe essere successivo. Dovresti comunque esserne diventato cosciente abbastanza presto per spiegare perché hai incluso quel contenuto in un successivo atto linguistico di memoria – assumendo che per definizione ogni oggetto nel tuo mondo eterofenomenologico sia un oggetto nella tua coscienza. Ciò fissa il momento *finale* in cui il contenuto «è diventato cosciente». Ma, come abbiamo visto, se questi due momenti determinano un periodo di alcune centinaia di millisecondi all'interno del quale la coscienza dell'oggetto deve verificarsi, e se ci sono diversi oggetti che devono presentarsi in quella finestra temporale (il punto rosso e quello verde; la donna dai capelli lunghi con e senza occhiali), non c'è modo di usare i tuoi *resoconti* per ordinare gli eventi che vengono rappresentati nella coscienza.

I tuoi resoconti verbali retrospettivi devono essere neutrali rispetto alle due presunte possibilità, ma gli scienziati non potrebbero trovare altri dati da usare? Potrebbero farlo se ci

fosse una buona ragione per ritenere che un qualche comportamento (esterno o interno) sia un buon segnale della coscienza. Ma questo è proprio ciò che la ragione ci nega. I due psicologi concordano nel ritenere che qualunque reazione comportamentale a un contenuto *potrebbe* essere meramente inconscia – se si esclude un successivo atto verbale. Nel modello staliniano c'è un'inconscia pressione del pulsante (e perché no?). Entrambi gli psicologi sono anche d'accordo che ci potrebbe essere una esperienza cosciente che non lascia effetti comportamentali. Nel modello orwelliano c'è una coscienza momentanea di un punto rosso immobile che non lascia traccia su nessuna reazione successiva (e perché no?).

Entrambi i modelli possono spiegare accuratamente *tutti* i dati – non solo i dati di cui già disponiamo, ma tutti quelli che possiamo immaginare di avere in futuro. Entrambi spiegano i resoconti verbali: una teoria dice che sono innocentemente errati, mentre l'altra dice che sono resoconti accurati di esperienze erronee. Inoltre, possiamo supporre, i due psicologi condividono *esattamente* la stessa teoria su ciò che accade nel tuo cervello; sono esattamente d'accordo su dove e quando nel cervello il contenuto erroneo è entrato nella catena causale; non sono d'accordo solo sul modo di caratterizzare quel luogo: uno lo ritiene pre-esperienziale, l'altro post-esperienziale. Danno la stessa spiegazione degli effetti non verbali, con una leggera differenza: uno dice che sono il risultato di contenuti inconsciamente discriminati, mentre l'altro dice che sono il risultato di contenuti consciamente discriminati ma dimenticati. Infine, entrambi rendono giustizia ai dati soggettivi – ogni cosa ottenibile dalla prospettiva in prima persona – perché si trovano d'accordo perfino su come tutto ciò dovrebbe «sembrare» ai soggetti: i soggetti dovrebbero essere incapaci di riconoscere la differenza tra «esperienze mal-acquisite» e «esperienze istantaneamente mal-ricordate».

Così, malgrado le apparenze, c'è solo una differenza verbale tra le due teorie (per una diagnosi simile, vedi Reingold e Merikle, 1990). Le due teorie raccontano esattamente la stessa storia tranne che per il luogo dove pongono il mitico Grande Spartiacque, un punto nel tempo (e quindi un punto nello spazio) la cui *esatta* posizione non può essere definita neanche con l'aiuto dei soggetti e che è neutrale anche rispetto a tutte le altre caratteristiche delle due teorie. Questa è una differenza che non fa differenza.

Considera un'analogia contemporanea. Nel mondo dell'edi-

toria c'è una distinzione tradizionale e usualmente ben definita tra una revisione antecedente la pubblicazione e una correzione di «errata» posteriore alla pubblicazione. Nel mondo accademico di oggi, però, le cose sono state molto accelerate dalle comunicazioni elettroniche. Con l'avvento della videoscrittura, del desktop publishing e della posta elettronica, ora accade spesso che varie versioni differenti di uno stesso articolo siano simultaneamente in circolazione, con l'autore che apporta rapidamente delle correzioni in risposta ai commenti che riceve attraverso la posta elettronica. Fissare un momento della pubblicazione, e quindi chiamare una delle versioni dell'articolo il testo *canonico* – il testo da archiviare, da citare nelle bibliografie – diventa in un certo senso un fatto arbitrario. Spesso la maggior parte dei lettori ai quali l'autore pensa, dei quali si preoccupa, leggono solo una prima versione; la versione «pubblicata» è per loro inerte, viene semplicemente archiviata. Se sono gli effetti importanti quelli che ci interessano, allora la maggior parte, se non tutti gli effetti importanti di un articolo su una rivista specializzata si verificano prima della pubblicazione e sono il frutto di varie differenti versioni. Un tempo le cose erano differenti; un tempo quasi tutti gli effetti importanti di un articolo si verificavano *dopo* la sua apparizione su una rivista e a *causa* di una tale apparizione. Ora che si può vedere come tutti i vari candidati alla «svolta» della pubblicazione non siano più funzionalmente importanti, se ancora sentiamo il bisogno di una tale distinzione, dobbiamo decidere arbitrariamente cosa deve valere come la pubblicazione di un testo. Non c'è un vertice o un punto di svolta naturale nel cammino che va dalla bozza della prima versione a quella da depositare in archivio.

Similmente – e questa è l'implicazione fondamentale del modello delle Molteplici Versioni – se si vuole stabilire che qualche momento nell'elaborazione del cervello debba valere come il momento della coscienza, si deve sapere che questa scelta è arbitraria. Si può sempre «tracciare una linea» nel flusso elaborativo del cervello, ma non ci sono differenze funzionali tali da giustificare che tutti gli stadi e le revisioni precedenti vengano dichiarati degli aggiustamenti inconsci o preconsci; e che tutte le successive correzioni del contenuto (così come è rilevato nel ricordo) vengano dichiarate delle contaminazioni post-esperienziali della memoria. Su una scala ridotta la distinzione crolla.

La regola pratica dell'astronomo:
se non ne prendi nota,
non è accaduto.

CLIFFORD STOLL,
L'uovo del cuculo, 1989

Come insegna ogni libro sui giochi di prestigio, i trucchi migliori sono quelli che finiscono prima che il pubblico pensi che siano cominciati. A questo punto penserai probabilmente che io ne abbia appena eseguito uno ai tuoi danni. Ho sostenuto che a causa della diffusione spazio-temporale del punto di vista dell'osservatore nel cervello, tutte le prove che abbiamo o che potremo avere non riescono a distinguere la teoria orwelliana dell'esperienza cosciente da quella staliniana e quindi tra le due teorie *non c'è differenza*. Questo è una specie di operazionismo o verificazionismo: esclude la possibilità che ci siano dei bruti fatti materiali che siano irraggiungibili dalla scienza, anche se la scienza include l'eterofenomenologia. Inoltre, sembra davvero piuttosto ovvio che *esistano* tali fatti bruti – che la nostra immediata esperienza cosciente consista di tali fatti!

Sono d'accordo che sembra piuttosto ovvio; se non lo fosse, non avrei dovuto impegnarmi così a fondo in questo capitolo per mostrare che ciò che è ovvio è in realtà falso. Ciò che sembra io abbia escluso, deliberatamente, è qualcosa di analogo al deriso Teatro Cartesiano della Coscienza. Potrai sospettare che, con il pretesto di combattere il dualismo («Sbarazziamoci dei fantasmi!»), io abbia fatto scomparire (piuttosto letteralmente) qualcosa su cui Cartesio aveva in effetti ragione: c'è da qualche parte una sorta di luogo funzionale dove gli oggetti della fenomenologia vengono... *proiettati*.

È tempo di confrontarci con questo sospetto. Nelson Goodman solleva il problema quando dice che l'esperimento *phi* colorato di Paul Kolers «sembra lasciarci una scelta tra una teoria della costruzione retrospettiva e una fede nella chiaroveggenza» (Goodman, 1978, p. 83). Dovendo scartare la chiaroveggenza, cos'è allora esattamente una «costruzione retrospettiva»?

Se pensiamo che la percezione del primo lampo sia differita o conservata o ricordata, abbiamo ciò che io chiamo una teoria della costruzione retrospettiva – la teoria secondo la quale la costruzione che

viene percepita come verificantesi tra i due lampi in realtà non è realizzata se non dopo il secondo.

Per prima cosa, Goodman sembra vacillare tra una teoria staliniana (la percezione del primo lampo è differita) e una teoria orwelliana (la percezione del primo lampo è conservata o ricordata), ma ciò che appare molto più importante è che il revisore da lui postulato (sia esso orwelliano o staliniano) non riadatta semplicemente i giudizi, *costruisce* il materiale con cui *riempire* le lacune:

ogni spazio che intercorre nel tragitto tra i due lampi è riempito [...] con uno dei colori lampeggiati piuttosto che con colori intermedi successivi. (p. 85)

Goodman tralascia la possibilità che il cervello non debba in realtà preoccuparsi di «riempire» un bel nulla con una «costruzione» – perché nessuno sta guardando. Come il modello delle Molteplici Versioni afferma esplicitamente, una volta che una discriminazione è stata eseguita, essa non deve essere eseguita di nuovo; il cervello si adatta alla conclusione che è stata tratta, rendendo disponibile per la modulazione del comportamento successivo la nuova interpretazione dell'informazione.

Goodman prende in considerazione la teoria, che egli attribuisce a Van der Waals e Roelofs (1930), secondo la quale «il movimento intermedio viene prodotto retrospettivamente, viene costruito solo dopo che si è verificato il secondo lampo e viene *proiettato indietro nel tempo* [il corsivo è mio]» (pp. 73-74). Ciò suggerisce un'opinione staliniana con una singolare complicazione: il film, dopo essere stato composto, viene fatto scorrere su un magico proiettore il cui raggio luminoso in qualche modo viaggia indietro nel tempo sullo schermo della mente. Indipendentemente dal fatto se sia questa la giusta interpretazione da dare a quello che Van der Waals e Roelofs avevano in mente quando hanno proposto la «costruzione retrospettiva», questa magia è presumibilmente ciò che ha condotto Kolers (1972, p. 184) a rifiutare la loro ipotesi, insistendo che tutte le costruzioni debbono essere effettuate nel «tempo reale». Perché mai, allora, il cervello si dovrebbe preoccupare di «produrre» il «movimento intermedio»? Perché il cervello non potrebbe semplicemente *concludere che c'era* un movimento intermedio e inserire questa conclusione retrospettiva nel flusso dell'elaborazione? Questo non è sufficiente?

Alt! Qui è dove (l'eventuale) gioco di prestigio deve verificarsi. Dal punto di vista in terza persona, io ho postulato un soggetto, il soggetto eterofenomenologico, una specie di fittizio «a chiunque possa interessare» al quale, in effetti, noi dall'esterno attribuiamo correttamente la convinzione di aver avuto esperienza del movimento intermedio. Così è come dovrebbe sembrare a *questo* soggetto (che è solo una finzione teorica). Ma non c'è anche un *vero* soggetto, per il quale il cervello mette in scena lo spettacolo, riempiendo tutte le lacune? Questo è quello che Goodman sembra supporre quando parla del cervello che riempie tutti gli spazi del tragitto. Per il piacere di chi vengono eseguiti tutti questi cartoni animati? Per il pubblico nel Teatro Cartesiano. Ma *poiché non c'è un tale teatro, non c'è un tale pubblico.*

Il modello delle Molteplici Versioni è d'accordo con Goodman sul fatto che il cervello crea retrospettivamente il contenuto (il giudizio) che c'era un movimento intermedio e questo contenuto è poi disponibile per governare le attività e lascia il suo segno nella memoria. Ma il modello delle Molteplici Versioni ritiene inoltre che il cervello non si preoccupa di «costruire» alcuna rappresentazione che serve allo scopo di «riempire» le lacune. Ciò sarebbe uno sciupio di tempo e (possiamo dirlo?) di *vernice*. Il giudizio è *già* disponibile, così il cervello può occuparsi di altri compiti![5]

La «proiezione indietro nel tempo» di Goodman è un'espressione equivoca. Potrebbe significare qualcosa di modesto e difendibile, e cioè che un *riferimento a qualche tempo passato* venga incluso nel contenuto. In questa lezione assomiglierebbe a qualcosa come «questo romanzo ci riporta indietro all'antica Roma...», che nessuno vorrà interpretare in un modo metafisicamente stravagante e ritenere che il romanzo sia una specie di macchina per viaggiare nel tempo. Questa è la lezione coerente con le altre opinioni di Goodman, ma Kolers sembra attribuirgli un significato metafisicamente più radicale, come se ci fosse una qualche effettiva proiezione di qualcosa da un tempo all'altro.

Come vedremo nel prossimo capitolo, la confusione provocata da questa lettura radicale della «proiezione» ha frustrato le interpretazioni di altri fenomeni. La stessa curiosa metafisica una volta ossessionava le riflessioni sulla rappresentazione dello spazio. Ai tempi di Cartesio, Thomas Hobbes sembra aver pensato che dopo che la luce ha colpito gli occhi e prodotto una specie di movimento nel cervello, ciò induce qualcosa a rimbalzare in qualche modo nel mondo.

La causa della sensazione è il corpo esterno, o l'oggetto, che agisce sull'organo proprio a ciascun senso, o immediatamente, come nel caso del gusto e del tatto, o mediatamente, come nella vista, nell'udito e nell'odorato; questa azione, attraverso la mediazione dei nervi e delle altre fibre e membrane del corpo, viene trasmessa internamente al cervello e al cuore, e causa una resistenza, o reazione o sforzo del cuore per liberarsene; tale sforzo, poiché si dirige verso l'esterno, sembra qualcosa di esteriore. (*Leviatano*, Parte I, cap. 1, «Il senso»)

Dopo tutto, egli pensava, è lì che vediamo i colori – sulla superficie degli oggetti![6] In maniera simile si potrebbe supporre che quando inciampi, un segnale parte dal tuo dito e arriva ai «centri del dolore» del cervello, che poi «proiettano» il dolore *giù fino al dito a cui appartiene*. Dopo tutto, è lì che il dolore è sentito.

Ancora negli anni Cinquanta di questo secolo tale idea era presa così seriamente da spingere lo psicologo britannico J.R. Smythies a scrivere un accurato articolo che la demolisse.[7] La proiezione di cui parliamo in questi fenomeni *non* implica l'irradiazione di qualche effetto nello spazio fisico e io credo che nessuno oggigiorno lo pensi più. I neurofisiologi e gli psicologi, e anche i tecnici del suono che progettano sistemi di amplificazione stereofonica, parlano spesso di questo tipo di proiezione, comunque, e potremmo chiederci che cosa intendono esattamente se non qualcosa che implica la trasmissione fisica da un punto (o un tempo) ad un altro. Che cosa implica? Osserviamo attentamente un semplice caso.

Grazie alla dislocazione degli altoparlanti stereo e all'equilibrio del volume delle loro rispettive emissioni sonore, l'ascoltatore *proietta* il suono risultante del soprano in un punto a metà strada tra i due altoparlanti.

Cosa significa? Dobbiamo interpretarlo attentamente. Se gli altoparlanti diffondono il suono in una stanza vuota, non c'è nessuna proiezione. Se è presente un ascoltatore (un osservatore con buoni orecchi e buon cervello), si verifica la «proiezione», ma questo *non* significa che qualcosa è emesso dall'ascoltatore verso il punto a metà strada tra i due altoparlanti. Nessuna proprietà fisica di quel punto o di quelli circostanti viene modificata dalla presenza dell'ascoltatore. In breve, questo è quello che intendiamo quando diciamo che Smythies aveva ragione; non c'è nessuna proiezione nello spazio di alcuna proprietà visiva o auditiva. Che cosa accade allora? Beh, *sembra all'osservatore* che la voce del soprano provenga da quel punto.

Che cosa implica questo *sembrare ad un osservatore*? Se rispondiamo che implica «la proiezione da parte dell'osservatore della voce in quel punto dello spazio» ci ritroviamo di nuovo al punto di partenza, ovviamente, così la gente è tentata di introdurre qualcosa di nuovo, dicendo qualcosa del genere: «l'osservatore proietta la voce *nello spazio fenomenologico*». Questo sembra un progresso. Abbiamo negato che la proiezione sia nello spazio fisico e l'abbiamo spostata nello spazio fenomenologico.

Ora, che cos'è lo spazio fenomenologico? È uno spazio fisico nel cervello? È lo spazio del palcoscenico nel teatro della coscienza collocato nel cervello? Non letteralmente. Ma metaforicamente? Nel capitolo precedente abbiamo visto un modo per ricavare un senso da tali spazi metaforici, nell'esempio delle «immagini mentali» che Shakey manipolava. In un senso stretto ma metaforico, Shakey tracciava linee nello spazio, prestava attenzione a specifici punti nello spazio, traeva conclusioni su ciò che trovava in quei punti dello spazio. Ma lo spazio era solo uno spazio *logico*. Era come lo spazio della Londra di Sherlock Holmes, uno spazio di un mondo fittizio, ma un mondo fittizio sistematicamente ancorato a concreti eventi fisici nello spazio ordinario nel «cervello» di Shakey. Se prendiamo le frasi pronunciate da Shakey come espressioni delle sue «credenze», allora possiamo dire che era *lo spazio in cui Shakey credeva*, ma questo non lo trasformava in uno spazio reale, come la fede di qualcuno in Phaenomeno non renderebbe Phaenomeno reale. Sono entrambi degli oggetti meramente intenzionali.[8]

Così abbiamo *un* modo per rendere sensata l'idea dello spazio fenomenico – come spazio logico. Questo è uno spazio nel quale o all'interno del quale nulla viene letteralmente proiettato; le sue proprietà sono costituite semplicemente dalle credenze del soggetto (eterofenomenologico). Quando diciamo che l'ascoltatore proietta il suono in un punto in questo spazio, noi intendiamo *soltanto* che a lui sembra che il suono proviene da lì. Non è abbastanza? Stiamo forse lasciandoci sfuggire una dottrina «realista» dello spazio fenomenico, in cui il *reale sembrare* possa essere proiettato?

Oggigiorno ci siamo abituati, senza tanti problemi, alla distinzione tra la collocazione spaziale nel cervello del veicolo dell'esperienza e la collocazione «nello spazio esperienziale» di ciò di cui facciamo esperienza. In breve, distinguiamo il rappresentante dal rappresentato, il veicolo dal contenuto. Siamo diventati abbastanza sofisticati da riconoscere che i prodotti della percezione visiva non sono, letteralmente, fotografie nella

testa, anche se *ciò che rappresentano* è ciò che le fotografie rappresentano bene: la disposizione spaziale di varie proprietà visive. Dovremmo fare la stessa distinzione per il tempo: *quando* accade un'esperienza nel cervello deve essere distinto da quando sembra accadere. Infatti, come lo psicolinguista Ray Jackendoff ha suggerito, il punto da capire in questo caso è solo una semplice estensione del comune buon senso sull'esperienza dello spazio. La rappresentazione dello spazio nel cervello non deve sempre usare «spazio nel cervello» per rappresentare lo spazio, e la rappresentazione del tempo nel cervello non deve sempre usare «tempo nel cervello». Così infondato come il proiettore di diapositive spaziali che Smythies non riusciva a trovare nel cervello è il proiettore di film temporali che viene suggerito da una lettura radicale dell'espressione di Goodman «proiezione indietro nel tempo».

Perché la gente sente il bisogno di postulare un simile proiettore delle apparenze? Perché è incline a pensare che non sia sufficiente che le sale di montaggio nel cervello inseriscano semplicemente il contenuto nel flusso che conduce alla modulazione del comportamento e alla memoria? Forse perché vuole conservare la distinzione realtà/apparenza per la coscienza. Vuole resistere al diabolico operazionismo che dice che ciò che è accaduto (nella coscienza) è semplicemente ciò che ricordi sia accaduto. Il modello delle Molteplici Versioni stabilisce il «prenderne nota» nella memoria come criterio per la coscienza; questo è *il modo in cui* il «dato» viene «preso» – preso in un modo piuttosto che in un altro. Non c'è una realtà dell'esperienza cosciente indipendente dagli effetti dei vari veicoli del contenuto sull'azione successiva (e quindi, naturalmente, sulla memoria). Ciò sembra sinistramente simile al temuto operazionismo, e forse il Teatro Cartesiano della coscienza è nascostamente amato come il luogo dove accade *davvero* qualunque cosa avviene «nella coscienza», indipendentemente dal fatto che sia o non sia poi ricordato. Supponiamo che in mia presenza accada qualcosa che lascia la sua traccia su di me solo per «un milionesimo di secondo», come nell'epigramma di Ariel Dorfman. Cosa mai potrà significare dire che io ne sono stato cosciente in modo brevissimo e che non lascia effetti? Se ci fosse il Teatro Cartesiano da qualche parte, almeno si potrebbe dire che *lì il film è stato sicuramente mostrato* anche se non c'è nessuno che se ne ricorda. (E con questo!)

Il Teatro Cartesiano può rappresentare un'immagine confortante perché mantiene la distinzione realtà/apparenza nel

cuore della soggettività umana, ma questo oltre ad essere scientificamente immotivato è metafisicamente dubbio, perché crea la bizzarra categoria dell'oggettivamente soggettivo – il modo in cui le cose ti sembrano realmente e oggettivamente, anche se non ti sembra che a te sembrino così! (Smullyan, 1981). Alcuni pensatori odiano talmente il «verificazionismo» e l'«operazionismo» da negarlo anche nell'arena dove dà manifestamente buona prova di sé: il regno della soggettività. Ciò che Clifford Stoll chiama la regola pratica dell'astronomo è un sardonico commento sulle stravaganze della memoria e gli standard delle prove scientifiche, ma diventa la verità letterale se applicato a ciò che viene «scritto» nella memoria. Potremmo classificare il modello delle Molteplici Versioni, allora, come *operazionismo in prima persona*, perché rifiuta nettamente la possibilità in linea di principio che si possa avere coscienza di uno stimolo senza che il soggetto creda in questa coscienza.[9]

L'opposizione a questo operazionismo fa appello, come al solito, a possibili fatti che trascendono la conoscenza dei metodi di prova dell'operazionista, ma ora l'operazionista è il soggetto stesso, e così l'obiezione si ritorce contro chi l'avanza: «Solo perché non puoi dire, tramite i tuoi accessi privilegiati, se sei stato o no conscio di *x*, non significa che non lo sei stato. Forse sei stato cosciente di *x*, ma non puoi trovare nessuna prova per esserne sicuro!». Qualcuno, dopo averci riflettuto, potrebbe davvero voler dire ciò? Presunti fatti relativi alla coscienza, che navigano alla deriva fuori della portata sia degli osservatori «esterni» che di quelli «interni», sono davvero degli strani fatti.

L'idea è difficile da sradicare. Considera quanto ci appare naturale la frase: «Ho giudicato così, perché è così che mi sembrava»; con la quale siamo incoraggiati a pensare a due eventi o stati distinti: «il sembrare in una certa maniera» e il successivo (e conseguente) «giudicare che le cose stanno in quella maniera». Si potrebbe pensare che il problema che si presenta con la spiegazione che il modello delle Molteplici Versioni dà del fenomeno *phi* colorato, per esempio, sia che sebbene includa il giudizio del soggetto che si è verificato un movimento intermedio, non include – anzi ne nega esplicitamente l'esistenza – alcun evento che potrebbe essere chiamato il «sembrare che si sia verificato un movimento intermedio», sul quale il giudizio è «basato». Da qualche parte ci devono essere delle «prove presentate», anche soltanto in un processo

farsa staliniano, cosicché il giudizio si può ritenere causato da, o fondato su, quelle prove.

Alcune persone ritengono che questo modo intuitivo di pensare sia sorretto dalla fenomenologia. Hanno l'impressione di osservarsi davvero mentre giudicano che le cose stanno così come *risultato* del fatto che le cose sembrano a loro così. Nessuno ha mai osservato una cosa del genere «nella propria fenomenologia» poiché un fatto del genere sulla causalità sarebbe inosservabile (come Hume ha sottolineato molto tempo fa).[10]

Domanda a un soggetto in un esperimento sul *phi* colorato: hai giudicato che il punto rosso si muovesse verso destra e cambiasse colore perché ti sembrava che fosse così o ti sembrava che si muovesse perché tale era il tuo giudizio? Supponiamo che il soggetto dia una risposta «sofisticata»:

Lo so bene che non c'era nessun punto luminoso che si muoveva *davvero* – si tratta solo di movimenti apparenti, dopo tutto – ma so anche che il punto *sembrava* muoversi, così, oltre al mio giudizio sul punto che sembrava muoversi, c'è un evento *sul quale* è stato formulato il mio giudizio: il «sembrare muoversi» del punto. Non c'era nessun reale movimento, così ci deve essere stato un reale «sembrare muoversi» che costituisce la base del mio giudizio.

Forse il Teatro Cartesiano è popolare perché è il luogo dove i «sembrare» possono verificarsi in aggiunta ai giudizi. Ma il sofisticato argomento appena presentato è fallace. Postulare un «reale sembrare» *oltre* al giudicare o «parlare» espresso dal resoconto del soggetto equivale a moltiplicare le entità senza necessità. Peggio, equivale a una moltiplicazione impossibile delle entità; la specie di presentazione interna in cui il reale sembrare avviene è un disperato stratagemma metafisico, un modo per cercare di avere la botte piena e la moglie ubriaca, soprattutto perché quelli che sono propensi a parlare così insistono ansiosamente che la presentazione interna *non* avviene in nessuna specie di spazio dualistico e misterioso pervaso da etere cartesiano. Quando si abbandona il dualismo cartesiano, si deve abbandonare davvero anche lo spettacolo che doveva svolgersi nel Teatro Cartesiano, e con esso il suo pubblico, giacché né lo spettacolo né il pubblico possono essere trovati nel cervello, e il cervello è l'unico posto dove è possibile cercarli.

Torniamo al modello delle Molteplici Versioni, cercando di ampliare la sua descrizione e prendendo in esame un po' più dettagliatamente la situazione nel cervello che ne costituisce le fondamenta: per semplicità, mi concentrerò su quello che accade nel cervello durante l'esperienza visiva. Più tardi potremo estendere la spiegazione ad altri fenomeni.

Gli stimoli visivi provocano serie di eventi nella corteccia che conducono gradualmente a varie discriminazioni di specificità sempre maggiore. In tempi e luoghi differenti, varie «decisioni» o «giudizi» vengono formulati; più letteralmente, parti del cervello vengono messe in stati che discriminano caratteristiche differenti, per esempio prima il semplice insorgere degli stimoli, poi la collocazione, poi la forma, più tardi il colore (su percorsi differenti), più tardi ancora il movimento (apparente) ed infine il riconoscimento dell'oggetto. Questi stati discriminativi localizzati trasmettono i loro effetti in altre parti, contribuendo ad ulteriori discriminazioni, e così via (Van Essen, 1979; Allman, Meizin e McGuinness, 1985; Livingstone e Hubel, 1987; Zeki e Shipp, 1988). La domanda ingenua ma naturale che ci viene da chiedere è: dove confluiscono tutte queste informazioni? La risposta è: da nessuna parte. Alcuni di questi stati distribuiti e ricchi di contenuto muoiono presto senza lasciare tracce ulteriori. Altri lasciano invece delle tracce, sui successivi resoconti verbali dell'esperienza e della memoria, sulla «prontezza semantica» e su altri aspetti dell'insieme percettivo, sullo stato emotivo, sulle propensioni comportamentali, e così via. Alcuni di questi effetti – per esempio, le influenze sui successivi resoconti verbali – sono come minimo sintomatici della coscienza. Ma non c'è un unico luogo del cervello attraverso il quale tutte queste successioni causali debbono passare per depositare il loro contenuto «nella coscienza».

Non appena una di queste discriminazioni è stata compiuta, diventa disponibile per produrre un qualche comportamento, per esempio la pressione di un pulsante (o un sorriso o un commento), o per modulare un qualche stato informazionale interno. Per esempio, la discriminazione di una fotografia di un pesce potrebbe creare un «insieme percettivo» – facendo in modo che sia temporaneamente più facile vedere pesci (o anche soltanto animali in generale) in altre fotografie – o potrebbe attivare un particolare dominio semantico, facendo in modo che sia più probabile intendere la parola «amo» come uno stru-

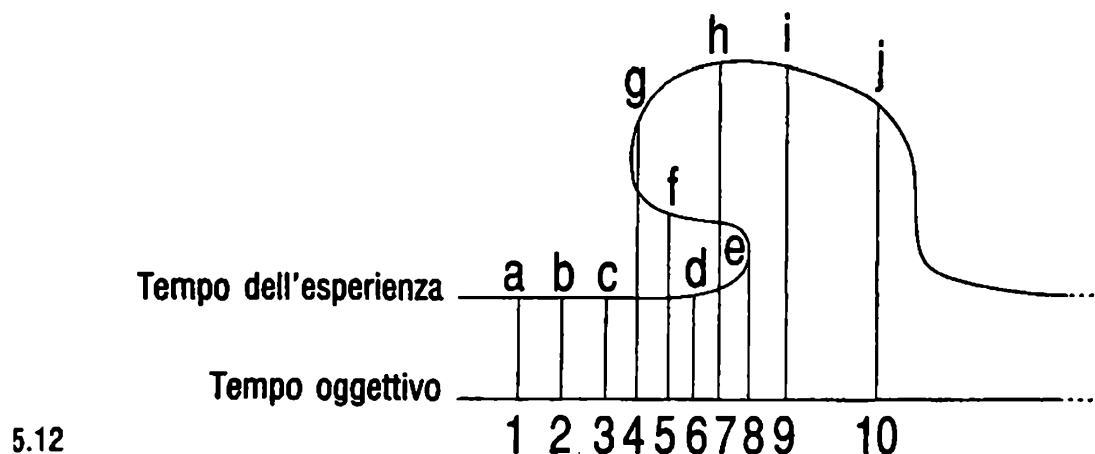
mento da pesca, piuttosto che come una mia propensione affettiva. Come abbiamo già notato, questa elaborazione su «piste multiple» prende varie centinaia di millisecondi, durante i quali possono verificarsi, in diverso ordine, varie aggiunte, fusioni, correzioni e riscritture del contenuto. Ciò conduce, nel corso del tempo, a qualcosa che *è abbastanza simile* a un flusso o una sequenza narrativa, che, si può pensare, è soggetta ad una continua e incessante revisione editoriale ad opera di molteplici processi distribuiti nel cervello che continueranno indefinitamente nel futuro. I contenuti sorgono, vengono revisionati, contribuiscono all'interpretazione di altri contenuti o alla modulazione del comportamento (verbale o d'altro tipo), e in questo processo lasciano le loro tracce nella memoria, infine degenerano o vengono incorporati o riscritti in contenuti successivi, del tutto o in parte. Questo groviglio di contenuti è simile solo in parte ad una narrazione per la sua molteplicità; in ogni momento ci sono molteplici versioni di frammenti di narrazione, a vari livelli di elaborazione, in vari luoghi del cervello. Mentre alcuni contenuti presenti in queste versioni daranno il loro effimero contributo e poi svaniranno senza produrre ulteriori effetti – e alcuni non apporteranno nessun contributo – altri continueranno a giocare molteplici ruoli nella ulteriore modulazione dello stato interno e del comportamento, e alcuni perdureranno così a lungo da fare la loro comparsa nelle dichiarazioni stampa rilasciate nella forma di vari comportamenti verbali.

Sondare questo flusso a vari intervalli produce effetti differenti, precipita differenti narrazioni – e queste *sono* narrazioni: singole versioni di una parte del «flusso di coscienza». Se si aspetta troppo a sondare, forse non ci sarà più una narrazione; se si sonda «troppo presto», si potranno ottenere dati che ci diranno quanto presto una particolare discriminazione è stata realizzata nel flusso, ma al costo di sconvolgere la normale progressione del flusso.

C'è un «tempo ottimale per il sondaggio»? Assumendo plausibilmente che dopo un po' queste narrazioni degenerano progressivamente, o perché i dettagli svaniscono, o perché vengono abbellite (ciò che avrei dovuto dire alla festa tende a diventare quello che ho detto alla festa), si potrebbe giustificare un sondaggio che avviene immediatamente dopo la sequenza di stimoli che ci interessa. Ma nello stesso tempo si vuole evitare di interferire nei fenomeni con un sondaggio prematuro. Poiché le percezioni si trasformano impercettibilmente in memorie e le

interpretazioni «immediate» si trasformano impercettibilmente in ricostruzioni razionali, non esiste un singolo punto di svolta adatto a tutti i contesti su cui dirigere la propria sonda.

Quello di cui siamo esattamente coscienti in uno specifico lasso di tempo non è definito indipendentemente dai sondaggi che usiamo per precipitare una narrazione su quel periodo. Poiché queste narrazioni sono sotto continua revisione, non esiste una singola narrazione che possa contare come la versione canonica, la «prima edizione» in cui sono presentati, una volta per tutte, gli eventi che si sono verificati nel flusso della coscienza del soggetto e che possa servire come parametro per valutare tutte le possibili deviazioni come corruzioni del testo. Ma ogni narrazione (o frammento di narrazione) che viene precipitata produce una «linea temporale», una sequenza soggettiva di eventi dal punto di vista dell'osservatore, che può essere confrontata con altre linee temporali, in particolare con la sequenza oggettiva di eventi che si verificano nel cervello dell'osservatore. Come abbiamo visto, queste due linee temporali *potrebbero non* coincidere in un allineamento ortogonale (potrebbero non collimare): anche se la discriminazione (errata) del *«rosso che diventa verde»* si è verificata nel cervello *dopo* la discriminazione del *punto verde*, la *sequenza soggettiva* o *narrativa* è, naturalmente, *punto rosso, poi «rosso che diventa verde», e infine punto verde.* Così all'interno della diffusione temporale del punto di vista del soggetto, ci possono essere ordini differenti che producono arricciamenti.

5.12

Non c'è nulla di stravagante o di provocatorio da un punto di vista metafisico in questa imprecisione dell'allineamento.[11] Non è più misterioso o contrario alla causalità del rendersi conto che le singole scene in un film spesso non sono girate in ordine cronologico o che quando leggi la frase: «Mario è

arrivato alla festa prima di Silvia, ma Gianna è arrivata prima di loro», tu vieni a sapere dell'arrivo di Mario prima di sapere del precedente arrivo di Gianna. Lo spazio e il tempo del rappresentante costituiscono una cornice di riferimento; lo spazio e il tempo di ciò che il rappresentante rappresenta ne costituiscono un'altra. Ma questo fatto metafisicamente innocuo sta tuttavia alla base di una fondamentale categoria metafisica: quando una porzione del mondo arriva in questa maniera a comporre un groviglio di narrazioni, quella porzione del mondo è un osservatore. Questo è quello che ci vuole perché ci sia un osservatore nel mondo, un qualcosa tale che si prova qualcosa a essere lui.

Questo è un abbozzo grossolano del mio modello alternativo. C'è ancora bisogno di chiarire bene le differenze che intercorrono tra questo modello e quello del Teatro Cartesiano, e questo verrà fatto mostrando come affronta alcuni fenomeni particolari. Nel prossimo capitolo, metteremo questo modello al lavoro su alcuni casi difficili, ma prima diamo un'occhiata rapida a qualche esempio più banale e familiare che spesso viene discusso dai filosofi.

Probabilmente ti è capitato di guidare per chilometri e chilometri assorto in una conversazione (o in un silenzioso soliloquio) per poi scoprire che non ti ricordi assolutamente della strada, del traffico, del tuo guidare, ecc. È come se avesse guidato qualcun altro. Molti pensatori (incluso me stesso, devo ammetterlo – Dennett, 1969, pp. 159 e seguenti) hanno voluto vedere in questo un caso esemplare di «percezione inconscia e azione intelligente». Ma eri *veramente* inconscio in quel momento di tutte le macchine che hai superato, dei semafori e delle curve che hai incontrato? Stavi prestando attenzione ad altro, ma sicuramente *se fossi stato sondato* su ciò che avevi *appena visto* in vari momenti del viaggio, avresti avuto almeno dei dettagli sommari da riferire. Il fenomeno della «guida inconscia» dovrebbe essere considerato piuttosto come un esempio di «coscienza che rotola» con rapida perdita di memoria.

Sei costantemente cosciente del ticchettio dell'orologio? Se improvvisamente si ferma, te ne accorgi e puoi dire immediatamente che cosa si è fermato; il ticchettio «di cui non eri cosciente» fino al momento in cui si è fermato e «di cui non saresti mai stato cosciente» se non si fosse fermato è ora chiaramente presente nella tua coscienza. Un caso ancor più sorprendente si verifica quando siamo in grado di contare, retrospettivamente nell'esperienza ricordata, i rintocchi di un

pendolo di cui abbiamo notato il suono solo dopo quattro o cinque rintocchi. Ma come è possibile che ci *ricordiamo* così chiaramente di *aver ascoltato* qualcosa di cui non siamo stati mai consci? La domanda tradisce un coinvolgimento con il modello cartesiano; non ci sono fatti stabiliti nel flusso della coscienza indipendentemente da sondaggi particolari.

6

IL TEMPO E L'ESPERIENZA

> Certo, io posso dire: le mie rappresentazioni si susseguono; ma ciò significa solamente: noi abbiamo coscienza di esse come in una serie temporale, cioè secondo la forma del senso interno.
>
> IMMANUEL KANT, *Critica della Ragion Pura*, 1781.

Nel capitolo precedente, abbiamo visto in linee generali come il modello delle Molteplici Versioni dissolve il problema della «proiezione indietro nel tempo», ma abbiamo sorvolato su alcune importanti complicazioni. In questo capitolo, ci occuperemo più approfonditamente di questi problemi e ci spingeremo in un territorio più infido: esamineremo e risolveremo svariate controversie che sono sorte tra psicologi e neuroscienziati a proposito della corretta interpretazione di alcuni famosi e problematici esperimenti. Ritengo che sia possibile comprendere il resto del libro senza seguire tutti gli argomenti di questo capitolo, quindi esso può essere saltato o soltanto scorso; malgrado ciò ho provato a mettere gli argomenti in una forma comprensibile anche per i non specialisti perché ci sono almeno sei buone ragioni per avventurarci negli aspetti tecnici.

1. Il mio abbozzo del modello delle Molteplici Versioni è in molti punti ancora oscuro e vederlo di nuovo in azione potrà servire a chiarire le idee sulla sua struttura.

2. I dubbi che eventualmente ancora rimangono sulla differenza tra il modello delle Molteplici Versioni, come teoria empirica, e il tradizionale Teatro Cartesiano possono essere dissolti osservando varie collisioni frontali tra le due teorie.

3. Se pensi che io stia attaccando un fantoccio, potrai scoprire come alcuni esperti si vanno a cacciare in un inestricabile groviglio perché, malgrado loro, sono autentici materialisti cartesiani.

4. Se ti viene il sospetto che io abbia basato il modello su un singolo fenomeno accuratamente scelto, il *phi* colorato di Kolers, potrai vedere come vari fenomeni molto differenti

ricevono una spiegazione migliore grazie alle Molteplici Versioni.

5. Molti dei famosi esperimenti che esamineremo sono stati presentati da *alcuni* importanti esperti come la confutazione del tipo di teoria materialistica tradizionale che io sto presentando, cosicché, se deve esserci una sfida *scientifica* alla mia spiegazione della coscienza, essa avviene su un terreno scelto dai miei avversari.

6. Infine, i fenomeni in questione sono affascinanti e vale la pena affaticarsi un po' per conoscerli meglio.[1]

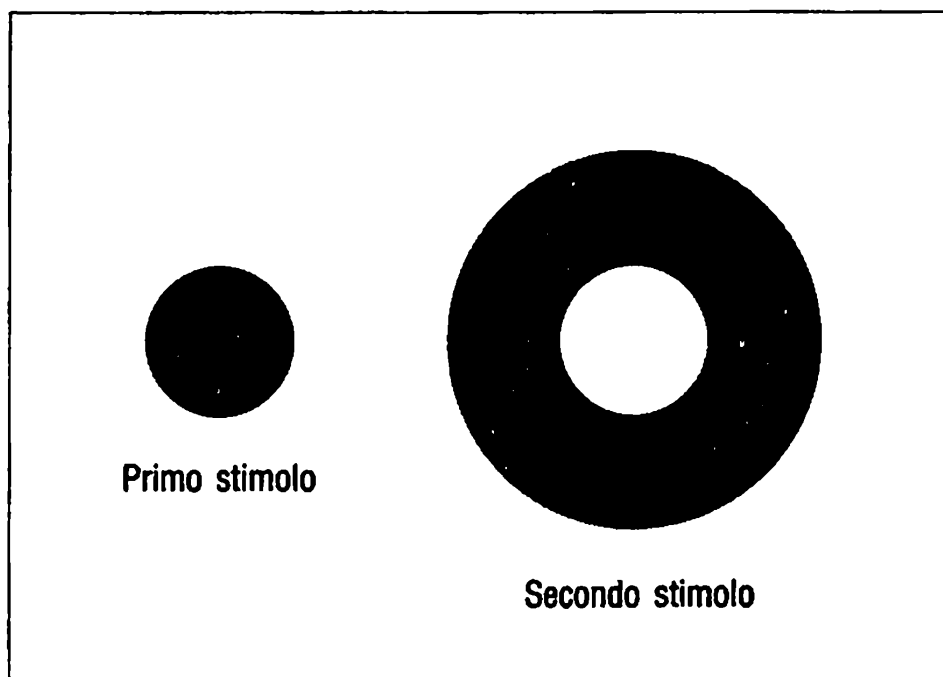
1. *Momenti fluttuanti e conigli saltellanti*

Una condizione normalmente sufficiente, ma non necessaria, per avere un'esperienza è un successivo resoconto verbale; questo è il caso fondamentale attorno al quale ruotano tutti i fenomeni enigmatici. Supponiamo che, benché il tuo cervello abbia già registrato – risposto a – alcuni aspetti di un evento, un qualcosa intervenga tra questa risposta interna e un'occasione successiva di un resoconto verbale. Se non c'è stato il tempo o l'opportunità per una iniziale risposta esplicita di qualche tipo e se gli eventi che intervengono impediscono delle successive risposte esplicite (verbali o d'altro genere) che facciano in qualche modo riferimento a qualche aspetto del primo evento, ci troviamo di fronte ad una situazione problematica: quegli aspetti non sono mai stati percepiti consciamente o sono stati rapidamente dimenticati?

Molti esperimenti hanno misurato «l'ampiezza dell'apprensione». In un test per l'ampiezza della memoria acustica, il soggetto ascolta una registrazione che elenca rapidamente una serie di termini non correlati (diciamo, quattro al secondo) e poi deve identificarli. Non si può rispondere fino a che l'evento acustico non sia terminato. Alcuni termini vengono identificati, ma non tutti. Eppure il soggetto ha potuto ascoltarli tutti in modo ugualmente chiaro e distinto. La domanda che ci viene spontanea è: di che cosa è stato esattamente cosciente il soggetto? Non c'è alcun dubbio che tutte le informazioni sul nastro sono state elaborate dal suo sistema auditivo, ma i segni distintivi dei termini che non sono stati successivamente menzionati nel resoconto verbale sono arrivati fino alla coscienza o

sono stati registrati solo inconsciamente? *Sembravano* lì, nella coscienza, ma vi erano *realmente*?

In un altro esempio sperimentale, al soggetto viene mostrata rapidamente una diapositiva nella quale sono disegnate delle lettere. (Ciò viene realizzato tramite un *tachistoscopio*, un congegno che può presentare un particolare stimolo visivo con la voluta luminosità e durata – talvolta anche solo 5 msec, altre volte 500 msec o più.) Nel successivo resoconto verbale, il soggetto riconosce solo alcune lettere, ma ha visto sicuramente anche le altre; egli può essere convinto che ce ne fossero altre, sapere perfino quante fossero, avere l'impressione che fossero ben delineate e distinte; eppure non può identificarle. Sono state rapidamente dimenticate o non sono mai state consciamente percepite?

Il fenomeno molto studiato del *metacontrasto* (Fehrer e Raab, 1962) mette bene in evidenza l'aspetto centrale del modello delle Molteplici Versioni. (Per un'analisi di fenomeni simili, vedi Breitmeyer, 1984.) Se uno stimolo viene proiettato velocemente su uno schermo (diciamo, per 30 msec – suppergiù quanto un unico fotogramma televisivo) per poi essere seguito immediatamente dopo da un secondo stimolo «occultante», i soggetti *riferiscono* di aver visto solo il secondo stimolo. Il primo stimolo potrebbe essere un disco colorato e il secondo un anello colorato che circonda perfettamente lo spazio in cui si trovava il disco.

6.1

Anche tu, al posto del soggetto, giureresti di aver visto un solo stimolo: l'anello. Nella letteratura psicologica, la spiega-

zione standard di tali fenomeni è staliniana: il secondo stimolo *impedisce in qualche modo l'esperienza cosciente* del primo stimolo. In altre parole, interferisce nel viaggio che il primo stimolo compie verso la coscienza. Chiamati a indovinare se c'erano uno o due stimoli, i soggetti possono tuttavia ottenere dei risultati migliori di quelli ottenuti con risposte casuali. Ciò mostra di nuovo – dice il teorico staliniano – che gli stimoli possono avere la loro influenza su di noi senza che ne siamo coscienti. Il primo stimolo non si è mai presentato sul palcoscenico della coscienza, ma ha avuto i suoi effetti (quali che siano) in modo inconscio. Possiamo opporci a questa spiegazione del metacontrasto proponendo un'alternativa orwelliana: in realtà i soggetti sono stati coscienti del primo stimolo (il che spiega la loro capacità di indovinare correttamente), ma la loro memoria di questa esperienza cosciente è stata *quasi del tutto* cancellata dal secondo stimolo (il che spiega perché negano di averlo visto, malgrado la loro rivelatrice capacità d'indovinare). Ne risulta un'impasse, una difficoltà per entrambi i contendenti, poiché nessuno può escogitare un qualche esperimento cruciale che *possa* risolvere la disputa.

Ecco come il modello delle Molteplici Versioni tratta il fenomeno del metacontrasto. Quando molte cose avvengono rapidamente, il cervello può fare delle presupposizioni semplificanti. Il contorno esterno del disco si trasforma rapidamente nel contorno interno dell'anello. Il cervello, informato all'inizio solo del fatto che qualcosa stava accadendo (qualcosa dal contorno circolare in un determinato luogo), rapidamente riceve una conferma che c'è in effetti un anello, con bordo interno ed esterno. Senza attendere ulteriori indizi sulla presenza di un disco, il cervello arriva alla non troppo azzardata conclusione che c'era solo un anello. Dovremmo insistere nel dire che il disco è entrato nella coscienza perché sarebbe stato riferito *se l'anello non fosse sopraggiunto*? Ciò significherebbe commettere l'errore di supporre che sia possibile «bloccare lo scorrimento» del film nel Teatro Cartesiano e assicurarsi che il fotogramma con il disco sia davvero arrivato nel teatro prima che eventi successivi ne abbiano cancellato la memoria. Il modello delle Molteplici Versioni riconosce che l'informazione riguardo al disco si è trovata per un breve periodo in una posizione funzionale che ha contribuito ad un successivo resoconto, ma questo stato è svanito; non c'è ragione per sostenere che questo stato fosse all'interno dell'incantata sfera della coscienza fino alla sua espulsione ad opera di una «riscrit-

tura», né per sostenere, al contrario, che non ha mai raggiunto questo stato privilegiato. Le versioni composte in vari tempi e luoghi del cervello sono state più tardi ritirate dalla circolazione, sostituite da versioni riviste, ma nessuna di esse può essere scelta come quella che definisce il contenuto della *coscienza*.

Un esempio ancor più sorprendente di questa capacità di revisione ci viene offerto dal fenomeno del *coniglio cutaneo*. Gli psicologi Frank Geldard e Carl Sherrick hanno descritto gli esperimenti originali nel 1972 (vedi anche Geldard, 1977; Geldard e Sherrick, 1983, 1986). Il soggetto viene invitato a tenere il braccio comodamente appoggiato su di un tavolo, e poi gli vengono applicati dei congegni meccanici in grado di somministrare piccoli colpi ad una distanza di circa dieci centimetri l'uno dall'altro. Una serie di colpetti ritmati vengono dati da tali congegni, ad esempio cinque al polso seguiti da due al gomito e poi altri tre sul bicipite. Tra un colpo e l'altro c'è un intervallo che varia dai 50 ai 200 msec; e quindi una serie di colpi può durare meno di un secondo o anche due o tre secondi. L'effetto sorprendente è che al soggetto questi colpi sembrano viaggiare lungo il braccio in una successione regolare e tra punti equidistanti – come se un animaletto stesse saltellandogli sul braccio. Ora, la prima cosa che viene da domandare è *come fa il cervello a sapere* che dopo i cinque colpetti al polso ne arrivano altri vicino al gomito? Nell'esperienza dei soggetti la «partenza» dei colpetti dal polso avviene con il secondo colpetto, eppure in prove di controllo in cui i successivi colpetti al gomito non vengono dati, i soggetti sentono tutti e cinque i colpetti nel polso come ci si aspetta. Il cervello ovviamente non può «sapere» del colpetto nel gomito fino a che non si è verificato. Se sei ancora intrappolato nel Teatro Cartesiano, probabilmente dirai che il cervello differisce l'esperienza cosciente fino a che tutti i colpetti siano stati «ricevuti» da qualche stazione intermedia tra il braccio e la sede della coscienza (ovunque essa sia), e che questa stazione intermedia rivede i dati per adattarli ad una teoria del movimento e poi invia la versione revisionata alla coscienza. Ma il cervello differisce sempre la risposta ad un colpetto nel caso ce ne fossero altri? Se no, come fa a «sapere» quando differire?

Il modello delle Molteplici Versioni mostra che questa è una domanda mal formulata. Nel corso del tempo il cervello discrimina sia lo spostamento (lungo il braccio) sia il numero dei colpetti. E benché nella realtà fisica i colpetti siano raggruppati

in luoghi particolari, l'assunzione semplificante è che siano distribuiti regolarmente lungo l'estensione spazio-temporale dell'esperienza. Il cervello si affida a questa interpretazione parsimoniosa ma errata *dopo* che i colpetti siano registrati, naturalmente, e ciò ha l'effetto di spazzare via le interpretazioni precedenti (parziali) dei colpetti, ma alcuni effetti collaterali di quelle interpretazioni possono sopravvivere. Per esempio, supponiamo di chiedere ai soggetti di premere un pulsante se sentono *due colpi nel polso*; non sarebbe sorprendente che possano iniziare a premere il pulsante *prima* che i colpetti nell'avambraccio siano discriminati, prima di quei colpetti la cui discriminazione provoca l'errata interpretazione del saltellamento lungo il braccio.

Dobbiamo essere particolarmente attenti a non commettere l'errore di supporre che il contenuto ottenuto da questo sondaggio precoce costituisca il «primo capitolo» del contenuto che troveremo nella narrazione se dovessimo sondare lo stesso fenomeno più tardi. Ciò non fa altro che confondere due «spazi» differenti: lo spazio del rappresent*ante* e lo spazio rappresent*ato*. Questo è un errore così facile e diffuso che merita un paragrafo a sé stante.

2. *Come il cervello rappresenta il tempo*

Il materialismo cartesiano, la posizione che nessuno dichiara di abbracciare ma che influenza il modo di pensare di quasi tutti, suggerisce la seguente immagine sotterranea. Noi sappiamo che l'informazione si muove da un punto all'altro del cervello, essendo elaborata da vari meccanismi in varie aree. Le nostre intuizioni ci suggeriscono che il nostro flusso di coscienza consista di eventi che si verificano in sequenza, e che in ogni istante ogni elemento di quella sequenza possa essere classificato o come già avvenuto «nella coscienza» o come non ancora arrivato «lì». E se le cose stanno così, allora (sembra che) i veicoli del contenuto che si muovono, con il loro contenuto, attraverso il cervello devono essere simili a vagoni ferroviari sui binari; l'ordine in cui passano in un certo punto sarà l'ordine in cui «arrivano» al teatro della coscienza e (quindi) «diventano coscienti». Per verificare *in quale punto* del cervello si verifica la coscienza, basta individuare tutti i percorsi dei veicoli dell'informazione e vedere in quale punto stanno passando nell'istante in cui diventano coscienti.

Una riflessione sul compito fondamentale del cervello ci mostrerà cosa c'è di sbagliato in quest'immagine. Il cervello ha il compito di guidare il corpo che controlla in un mondo di condizioni mutevoli e di improvvise sorprese, così deve raccogliere informazioni dal mondo e usarle *rapidamente* per «produrre futuro» – estrapolare delle anticipazioni allo scopo di stare un passo più avanti del disastro (Dennett, 1984a, 1991b). Il cervello deve quindi rappresentare le proprietà temporali degli eventi nel mondo, e deve farlo in modo efficiente. I processi che sono responsabili dell'esecuzione di questo compito sono spazialmente distribuiti in un grande cervello senza un nodo centrale e la comunicazione tra le parti del cervello è relativamente lenta; gli impulsi nervosi elettrochimici viaggiano a una velocità migliaia di volte più lenta di quella della luce (o dei segnali attraverso i circuiti elettronici). Il cervello si trova quindi ad agire sotto una significativa pressione temporale. Spesso deve riuscire a modulare i suoi segnali in uscita alla luce di quelli in ingresso entro una finestra temporale che non lascia gioco per dilazioni. Dalla parte degli ingressi, ci sono compiti di analisi percettiva, come la percezione di un discorso, che sarebbero oltre i limiti fisici del meccanismo cerebrale se non utilizzasse ingegnose strategie anticipatorie che sfruttano le ridondanze nei segnali in ingresso. Un discorso normale si sviluppa ad una velocità di quattro o cinque sillabe per secondo, ma i meccanismi di analisi linguistica che abbiamo sviluppato sono così potenti che la gente può comprendere un «discorso compresso» – in cui le parole sono elettronicamente accelerate senza che il suono divenga simile ad uno squittio di scoiattolo – fino alla velocità di trenta sillabe per secondo. Dalla parte dei segnali in uscita, molti atti devono avvenire così velocemente e con un «lancio» così accurato che il cervello non ha il tempo per aggiustare i suoi segnali di controllo alla luce di qualche retroazione; atti come il suonare il pianoforte o scagliare accuratamente un sasso (Calvin, 1983, 1986) devono essere iniziati *balisticamente*. (Un evento balistico è un missile che viaggia *senza* guida; una volta scagliato la sua traiettoria non è più aggiustabile.)

Come fa, allora, il cervello a registrare tutte le informazioni temporali che gli servono? Consideriamo il problema seguente. Poiché la distanza tra la punta del piede e il cervello è molto più grande di quella tra il bacino e il cervello, o le spalle e il cervello, o la fronte e il cervello, degli stimoli applicati simultaneamente in questi differenti siti arriveranno al Quartier Gene-

rale in successione sfalsata, se la velocità con cui viaggiano è costante lungo tutto il percorso. Come fa il cervello, ci si potrebbe chiedere, ad «assicurare la simultaneità centrale delle rappresentazioni per degli stimoli simultanei applicati a punti a diversa distanza da esso»? Volendosi impegnare in una sorta di ingegneria congetturale alla rovescia, si potrebbe pensare che forse tutti i nervi afferenti sono come quei metri a molla arrotolati su se stessi; sono tutti della stessa lunghezza: i nervi dalla punta del piede sono completamente dipanati, quelli dalla fronte sono quasi completamente avvolti nel cervello; i segnali dalla fronte nell'ultimo tratto girano in tondo tante volte quante sono le spire dell'avvolgimento dilatorio e si presentano al Quartier Generale nello stesso identico istante dei segnali non ritardati del piede. O si potrebbe immaginare che i nervi diventino più sottili man mano che si allungano (come gli elastici o gli spaghetti fatti in casa) e che la velocità di conduzione vari con il diametro. (È così in effetti, ma nel senso contrario, purtroppo! Più le fibre sono spesse, più la velocità di conduzione è alta.) Questi sono modelli pittoreschi (ma ingenui) dei meccanismi necessari per risolvere questo problema, ma l'errore dal quale sorgono è quello di supporre che il cervello debba davvero risolvere *questo* problema. Il cervello non deve risolvere questo problema, per un'ovvia ragione ingegneristica: spreca del tempo prezioso se alle sue operazioni concede tutto il tempo necessario alla «ipotesi peggiore». Perché dei segnali importanti e vitali dalla fronte (per esempio) dovrebbero gironzolare nell'anticamera solo perché un giorno potrebbe presentarsi il caso in cui debbano convergere con altri segnali concomitanti provenienti dai piedi.[2]

I calcolatori digitali dipendono effettivamente da tali dilazioni per lasciare spazio all'ipotesi peggiore e assicurare la sincronia. Il meccanismo di un circuito addizionatore parallelo che trattiene le somme calcolate, e sta in ozio fino a che un impulso di sincronizzazione non le rilascia, presenta una stretta parentela con i nervi arrotolati che abbiamo immaginato. E i costruttori di supercomputer devono preoccuparsi che i fili di connessione tra le varie parti abbiano la stessa lunghezza, il che spesso richiede l'inclusione di giri supplementari dei fili. Ma i calcolatori digitali possono permettersi queste inefficienze locali perché la loro velocità è altissima. (In effetti, con un mercato che favorisce la competizione per ottenere dei calcolatori sempre più veloci, queste piccole inefficienze temporali vengono ora ripensate; la ragione principale per la quale non è possibile

disfarsi di tali inefficienze è che gli ingegneri non sanno come progettare dei sistemi di calcolatori completamente asincroni, non regolati cioè da un impulso temporale base.)

Imporre una sincronia di base sulle operazioni richiede dilazioni. Come ingegneri alla rovescia, possiamo immaginare che se esiste qualche modo efficiente con il quale il cervello può rappresentare le informazioni sul tempo di cui ha bisogno senza ricorrere a dilazioni, l'evoluzione l'avrà sicuramente «trovato». In realtà modi siffatti esistono, come possiamo illustrare con un incidente storico che mostra il fenomeno ingigantito sia nel tempo che nello spazio.

Si considerino le difficoltà di comunicazione che si presentavano ai governanti di un impero così esteso come quello britannico prima dell'avvento della radio e del telegrafo. Controllare un impero su scala mondiale dal quartier generale di Londra non era sempre possibile. L'incidente più noto è sicuramente quello della Battaglia di New Orleans, l'8 gennaio 1815, quindici giorni dopo la firma, in Belgio, dell'armistizio che metteva termine alla guerra del 1812, tra Stati Uniti e Gran Bretagna. Più di mille soldati britannici furono uccisi in questa inutile battaglia. Possiamo usare questo disastro per vedere come funzionava il sistema. Supponiamo che nel giorno 1 il trattato sia stato siglato in Belgio e la notizia inviata per mare e per terra in America, India, Africa, ecc. Il giorno 15 la battaglia viene combattuta a New Orleans e la notizia della disfatta inviata per mare e per terra in Inghilterra, India, ecc. Il giorno 20, troppo tardi, la notizia del trattato (e l'ordine di deporre le armi) arriva a New Orleans. Il giorno 35, supponiamo, la notizia della sconfitta arriva a Calcutta, ma la notizia del trattato non vi arriva prima del giorno 40 (viaggia via terra, più lentamente). Al comandante in capo delle truppe britanniche in India, la battaglia «sembrerebbe» avvenuta prima della firma del trattato – se non ci fosse l'abitudine di datare le lettere e permettere quindi al ricevente di fare le dovute correzioni.[3]

Questi funzionari sparpagliati per il mondo risolvevano la maggior parte dei loro problemi di comunicazione di informazioni cronologiche incorporando le rappresentazioni delle relative informazioni temporali nel *contenuto* dei loro segnali, cosicché il momento di arrivo dei segnali stessi era *strettamente irrilevante* rispetto all'informazione che trasportavano. Una data scritta all'inizio della lettera (o un timbro datato sulla busta) dà al ricevente l'informazione sufficiente per sapere quando è stata spedita, un'informazione che sopravvive a qualunque ritardo

nella consegna.[4] Questa distinzione tra il tempo rappresentato (dal timbro) e il tempo del rappresentante (il giorno in cui la lettera arriva) è un esempio della familiare distinzione tra contenuto e veicolo dell'informazione. Benché questa particolare soluzione non sia disponibile per i comunicatori del cervello (poiché non «conoscono la data» quando spediscono i loro messaggi), il principio generale della distinzione tra contenuto e veicolo nei modelli di elaborazione dell'informazione del cervello è più importante di quanto usualmente si ritenga.[5]

In generale dobbiamo distinguere tra le caratteristiche dei rappresent*anti* e le caratteristiche dei rappresent*ati*. Qualcuno può gridare a pieni polmoni «sommessamente, sulla punta dei piedi», esistono gigantografie di oggetti microscopici e la logica non viene offesa da un quadro a olio di un artista intento a fare uno schizzo a carboncino. La frase iniziale di una descrizione scritta di un uomo in piedi non deve descrivere la sua testa, né l'ultima i suoi piedi. Questo principio si applica anche, sebbene meno ovviamente, al tempo. Pronunciamo la frase «un intenso e fugace lampo rosso»; l'inizio è «un intenso» e la fine «rosso»; questi fragmenti dell'evento costituito dal nostro discorso non sono essi stessi delle rappresentazioni dell'avvio o della conclusione del fugace lampo rosso (per un'osservazione simile, vedi Efron, 1967, p. 714). Nessun evento nel sistema nervoso può avere una durata nulla (come non può avere un'estensione nulla), così ha un avvio e una conclusione separati da una certa quantità temporale. Se l'evento stesso *rappresenta* un evento nell'esperienza, allora l'evento che rappresenta deve anch'esso avere una durata non nulla: un avvio, un centro e una conclusione. Ma non c'è ragione di supporre che l'inizio del rappresentante rappresenti l'inizio del rappresentato.[6] Benché differenti attributi siano in effetti estrapolati da differenti strutture neurali a velocità differenti (per es. la localizzazione distinta dalla forma distinta dal colore) e benché riusciamo, se chiamati a farlo, a reagire alla presenza dei vari attributi singoli con latenze differenti, noi percepiamo eventi e non un gocciolio successivamente analizzato di elementi percettivi o di attributi.[7]

Un romanzo o un resoconto storico non deve essere composto nell'ordine che alla fine rappresenta – qualche volta l'autore comincia dalla fine e procede a ritroso. Inoltre i racconti possono contenere dei flashback, in cui l'ordine in cui gli eventi *rappresentati* si sono verificati non è dato dall'ordine in cui i

rappresentanti si verificano. Allo stesso modo, la rappresentazione cerebrale di *A prima di B* non deve essere realizzata da:

prima

una rappresentazione di A,

poi

una rappresentazione di B.

La frase «B dopo A» è un esempio di un veicolo (orale) che rappresenta la precedenza di A rispetto a B, e anche il cervello può avvalersi della stessa libertà di collocazione temporale. Ciò che importa al cervello non è necessariamente il momento in cui i singoli eventi rappresentanti si verificano nelle varie parti del cervello (fintanto che si verificano in tempo per controllare ciò che devono!), ma il loro *contenuto temporale*. L'importante è, cioè, che il cervello possa procedere a controllare gli eventi «sotto la presupposizione che A sia avvenuto prima di B», indipendentemente dal fatto che l'informazione A sia o non sia entrata nei relativi sistemi cerebrali e riconosciuta come tale prima dell'informazione B. (Ricordati del comandante in capo a Calcutta: prima è stato informato della battaglia e poi dell'armistizio, ma poiché può estrapolare da ciò l'informazione che l'armistizio è avvenuto prima, può agire di conseguenza. Egli deve *giudicare* che l'armistizio è avvenuto prima della battaglia; ma non deve anche mettere in scena una qualche «ricostruzione storica» dei fatti nella quale si osserverà ricevere le lettere nel «giusto» ordine.)

Alcuni pensatori hanno sostenuto, comunque, che il tempo è l'unica cosa che la mente o il cervello *devono* rappresentare «con il tempo stesso». Il filosofo Hugh Mellor, nel suo libro *Real Time* (1981, p. 8) espone questa tesi in modo chiaro e vigoroso:

Supponiamo per esempio che io veda un evento *x* precederne un altro, *x**. Io devo vedere prima *x* e poi *x**, il mio guardare *x* è in qualche modo ricordato nel mio guardare *x**. Cioè, il mio guardare *x* influenza il mio guardare *x**: è questo che mi fa vedere – giustamente o no – che *x* precede *x** invece del contrario. Ma vedere che *x* precede *x** significa vedere prima *x*. Quindi l'ordine causale delle mie percezioni di questi eventi, fissando il loro ordine temporale che

io percepisco, fissa l'ordine temporale delle percezioni stesse [...] Il fatto sorprendente [...] che va notato, è che le percezioni di un ordine temporale hanno bisogno di percezioni temporalmente ordinate. *Nessun'altra proprietà o relazione deve essere così incorporata nelle percezioni che la rivelano* (il corsivo è mio): le percezioni della forma e del colore, per esempio, non devono esse stesse avere la forma o il colore corrispondente.

Questo è falso, ma contiene nondimeno qualcosa di vero. Poiché la funzione fondamentale della rappresentazione cerebrale è quella di controllare il comportamento in tempo reale, le determinazioni temporali dei rappresentanti sono *in una certa misura* essenziali al loro compito, in due modi.

Primo, all'inizio di un processo percettivo, le determinazioni temporali possono essere *ciò che stabilisce il contenuto.* Si consideri come distinguiamo un punto luminoso che si muove da destra a sinistra da uno che si muove da sinistra a destra su uno schermo cinematografico. L'*unica* differenza tra questi due casi può essere l'ordine temporale in cui vengono proiettati i due (o più) fotogrammi. Se viene proiettato prima A e poi B, si vede il punto che si muove in una direzione; se prima B e poi A, si vede il punto muoversi in senso inverso. L'unica differenza tra gli stimoli che il cervello può utilizzare per effettuare questa discriminazione della direzione è l'ordine in cui si verificano. Questa discriminazione è allora logicamente basata sulla capacità del cervello di effettuare una discriminazione dell'ordine temporale di una certa acuità. Poiché i fotogrammi cinematografici sono di solito presentati alla velocità di ventiquattro al secondo, sappiamo che il sistema visivo è in grado di definire l'ordine tra stimoli che si verificano nello spazio di 50 msec. Ciò significa che le effettive proprietà temporali dei segnali – i loro tempi di insorgenza, la loro velocità nel sistema e quindi i loro tempi di arrivo – devono essere accuratamente controllate fino a che tale discriminazione non sia effettuata. Altrimenti, l'informazione sulla quale la discriminazione deve essere basata sarà persa o oscurata.

Su scala più ampia questo fenomeno si presenta talvolta alla partenza di una regata velica: prima *vedi* la barca superare la linea di partenza e poi *senti* il colpo di partenza, ma questo significa davvero che l'imbarcazione è partita troppo presto? Non è logicamente possibile dirlo fino a che non si siano calcolati i differenti tempi di trasmissione del suono e della luce dal luogo della partenza a quello in cui hai fatto la discriminazione. Una volta che un giudizio sia stato formulato (*tutto in regola* o *l'imbarcazione numero 7 è partita troppo presto*), questo

contenuto può essere comunicato ai contendenti con comodo, senza preoccuparsi di quanto veloce o quanto lontano *esso* debba andare per realizzare il suo compito.

Così la determinazione temporale di qualche rappresentazione è importante *fino a che* una discriminazione come *da sinistra a destra* (o *partita troppo presto*) non sia stata fatta, ma una volta che sia stata fatta, localmente, da qualche circuito della corteccia (o da qualche osservatore nella barca della giuria), il contenuto del giudizio può essere inviato, senza troppa premura, in tutti i luoghi del cervello dove questa informazione può essere utilizzata. Solo così possiamo spiegare il fatto altrimenti sorprendente che la gente in alcuni casi può essere incapace di formulare dei giudizi d'ordinamento temporale migliori della semplice casualità, sebbene non commetta nessun errore in altri giudizi che richiedono logicamente una maggiore capacità di discriminazione temporale (come quello della direzione di un movimento). Per formulare i giudizi di alta qualità vengono usati sistemi discriminatori specializzati (e soprattutto localizzati).

La seconda limitazione sulle determinazioni temporali è stata già notata incidentalmente: non è importante l'ordine in cui i rappresentanti si verificano fintanto che si verificano in tempo per contribuire a controllare l'appropriato comportamento. La funzione che un rappresentante è chiamato a svolgere può dipendere dal rispetto di una *scadenza*, che può essere considerata una proprietà temporale del veicolo che svolge il compito di rappresentante. Ciò è particolarmente evidente in un sistema così a corto di tempo come quello dell'ipotetico Scudo Spaziale. Il problema non è quello di come far rappresentare accuratamente a dei calcolatori un lancio di missili, ma quello di riuscire a realizzare tale rappresentazione nel breve tempo che si ha a disposizione per fare qualcosa. Il messaggio che un missile è stato lanciato alle 6.04.23.678 (ora di New York) può accuratamente rappresentare l'ora del lancio per sempre, ma la sua utilità può svanire dopo le 6.05. Per qualsiasi compito di controllo, allora, c'è una *finestra temporale di controllo* all'interno della quale i parametri temporali dei rappresentanti possono, in linea di principio, essere spostati a volontà.

Le scadenze che limitano tali finestre non sono fissate, ma dipendono dal compito. Se, invece di intercettare missili, stai scrivendo le tue memorie o rispondendo a domande nelle udienze sul caso Watergate (Neisser, 1981), puoi rintracciare in qualsiasi ordine le informazioni che ti servono per stabilire la sequenza degli eventi nella tua vita e puoi anche prenderti del

tempo per trarne le implicazioni. O, considerando un caso intermedio più vicino ai fenomeni che stiamo trattando, supponiamo che tu stia andando alla deriva su una barca e ti chieda se ti stai avvicinando o allontanando da quello scoglio pericoloso che vedi in lontananza. Supponiamo che *in questo momento* tu conosca l'attuale distanza dallo scoglio (hai appena misurato, per esempio, l'angolo che sottende nel tuo campo visivo); per rispondere alla tua domanda puoi aspettare un po' e poi misurare l'angolo di nuovo, *oppure*, se mezz'ora fa hai fatto un'istantanea Polaroid dello scoglio, potresti misurare l'angolo in quella vecchia fotografia, fare qualche conto e valutare la distanza di allora. Per formulare un giudizio sulla direzione nella quale sei trascinato devi calcolare due distanze: la distanza a mezzogiorno e la distanza alle 12.30, diciamo, ma è indifferente quale distanza viene calcolata per prima. Devi, in ogni modo, essere capace di fare i calcoli abbastanza velocemente da poterti mettere, eventualmente, ai remi prima che sia troppo tardi.

Così la rappresentazione che il cervello fa del tempo è ancorata al tempo stesso in due modi: la stessa determinazione temporale del rappresentante può essere ciò che fornisce le prove o determina il contenuto, e l'utilità della rappresentazione temporale delle cose può svanire se la rappresentazione non avviene in tempo per svolgere il suo ruolo. Credo che Mellor valuti positivamente entrambi questi fattori e che li avesse in mente quando ha fatto l'affermazione che ho citato, ma egli ha commesso l'errore naturale di pensare che la loro combinazione limiti *completamente* la rappresentazione del tempo, e che quindi l'ordine del rappresentante rappresenti *sempre* l'ordine nel contenuto. Secondo la sua spiegazione, non c'è spazio per una «diffusione» temporale, mentre io sto sostenendo che *deve* esserci diffusione temporale – su piccola scala – *perché* ci deve essere diffusione spaziale (su piccola scala) del punto di vista dell'osservatore.

Le cause devono precedere gli effetti. Questo principio fondamentale fa sì che le finestre temporali di controllo siano limitate da entrambi i lati: dal primo istante in cui le informazioni possono arrivare al sistema e dall'ultimo istante in cui le informazioni possono contribuire causalmente a controllare un particolare comportamento. Non abbiamo ancora visto *come* il cervello possa utilizzare il tempo disponibile in una finestra di controllo per selezionare l'informazione che riceve e trasformarla in una «narra-

zione» coerente che viene usata per dirigere le risposte del corpo.

In che modo, allora, i processi cerebrali possono inferire le proprietà temporali? Sistemi che si basano su «timbri postali» o «datari» non sono teoricamente impossibili, ma esiste un modo più economico, meno sicuro ma biologicamente più plausibile: tramite ciò che potremmo chiamare una *sistemazione sensibile al contenuto.* Un'utile analogia potrebbe essere quella di uno studio cinematografico in cui il sonoro viene «sincronizzato» con il film. I vari frammenti di nastro sonoro possono di per sé aver perso tutte le loro indicazioni temporali, cosicché non c'è un modo semplice e meccanico per metterli in perfetta concordanza con le immagini. Spostandoli, però, avanti e indietro lungo la pellicola e stando attenti alle convergenze, si arriverà senza troppe difficoltà a un «miglior abbinamento». Il ciak all'inizio di ogni ripresa – «Scena terza, seconda, motore, azione, CIAK!» – fornisce una doppia indicazione, sia visiva che auditiva, che una volta sincronizzata mette nella giusta posizione tutto il resto del nastro e della pellicola. Ma generalmente ci sono moltissimi punti di corrispondenza importanti e quindi questa corrispondenza convenzionale all'inizio di ogni ripresa è solo un'utile ridondanza. Per ottenere una buona corrispondenza ci si basa sul *contenuto* della pellicola e del nastro, ma non su un'analisi sofisticata del contenuto. Un montatore che non conoscesse il giapponese troverebbe difficile e noiosa la sincronizzazione di una colonna sonora e di una pellicola giapponesi, ma non impossibile. Inoltre, l'ordine temporale delle fasi del processo di abbinamento è indipendente dal contenuto del prodotto; il montatore può lavorare sulla terza scena prima di sistemare la seconda scena e in linea di principio potrebbe anche eseguire tutto il lavoro andando all'indietro.

Processi piuttosto «stupidi» possono eseguire degli assestamenti e spostamenti simili nel cervello. Per esempio, si possono trovare facilmente degli analoghi temporali del problema spaziale della computazione della profondità di stereogrammi a punti casuali (vedi figura 5.7). In linea di principio, allora, il cervello può risolvere alcuni dei suoi problemi di inferenza temporale tramite un processo del genere, ricavando i dati non dall'occhio destro e sinistro, ma da quelle fonti d'informazione che sono implicate in un processo che richiede un giudizio temporale. Due punti importanti seguono da ciò. Primo, tali inferenze temporali possono essere ottenute (tali discrimina-

zioni temporali possono essere fatte) comparando il *contenuto* (di basso livello) di varie strutture ordinate di dati, e questo processo in tempo reale non deve necessariamente avvenire nell'ordine temporale che il suo prodotto alla fine rappresenta. Secondo, una volta che un'inferenza temporale del genere è stata ottenuta, il che può avvenire *prima* che le caratteristiche di alto livello siano state ricavate da altri processi, essa non deve essere ricavata di nuovo! Non deve esserci una rappresentazione *successiva* in cui le caratteristiche di alto livello siano «presentate» in una sequenza in tempo reale a beneficio di un secondo giudicatore della sequenza. In altre parole, avendo ottenuto le inferenze da queste giustapposizioni di informazioni temporali, il cervello può continuare a rappresentare i risultati in qualsiasi formato che soddisfi i suoi bisogni e le sue risorse – non necessariamente nel formato in cui «il tempo è usato per rappresentare il tempo».

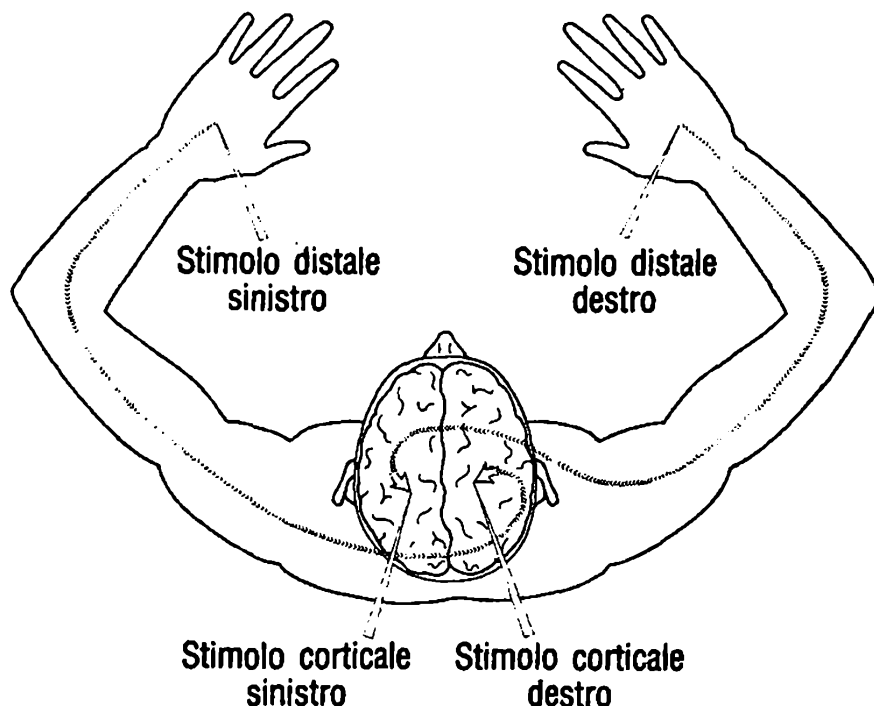
3. *Il caso del «riferimento all'indietro nel tempo» di Libet*

Abbiamo stabilito un modo in cui il cervello può svolgere il suo lavoro editoriale sulle informazioni temporali ignorando le effettive determinazioni temporali (il «tempo di arrivo») di alcune delle sue rappresentazioni, ma dobbiamo ricordarci ancora una volta delle forti pressioni temporali che condizionano questo lavoro. Procedendo a ritroso dalla scadenza, tutti i contenuti riferiti o altrimenti espressi nel successivo comportamento devono essere stati presenti (nel cervello, ma non necessariamente «nella coscienza») in tempo per contribuire causalmente a quel comportamento. Per esempio, se un soggetto in un esperimento *dice* «cane» in risposta ad uno stimolo visivo, possiamo procedere a ritroso da tale comportamento, che è stato chiaramente controllato da un processo che aveva il contenuto *cane* (a meno che quel soggetto non dica «cane» in risposta ad ogni stimolo o passi le sue giornate dicendo «cane, cane, cane...»). E dal momento che occorrono circa 100 msec per *iniziare* ad eseguire un'intenzione linguistica del genere (e circa altri 400 msec per completarla), possiamo essere abbastanza sicuri che il contenuto *cane* era presente, presumibilmente nelle aree linguistiche del cervello, 100 msec prima dell'inizio di quella espressione. Procedendo invece dall'altro capo, possiamo, di nuovo, determinare il momento in cui per la prima volta il sistema visivo può aver computato o estrapolato

dai dati in ingresso nella retina il contenuto *cane* e, forse, persino seguire la sua creazione e il suo successivo viaggio attraverso il sistema visivo verso le aree del linguaggio.

Un'anomalia davvero sorprendente (che ci farebbe effettivamente storcere il naso con sgomento) si avrebbe nel caso che il tempo trascorso tra lo stimolo *cane* e l'espressione «cane» fosse minore del tempo fisicamente necessario per la formazione di quel contenuto e la sua trasmissione nel sistema. Ma nessuna anomalia del genere è stata scoperta. Alcune sorprese sono emerse, comunque, dal confronto tra le due sequenze raffigurate nella figura 5.12. Quando cerchiamo di far coincidere la sequenza di eventi nel flusso di elaborazione oggettiva nel cervello con la sequenza di eventi soggettiva *così come è riportata dal soggetto successivamente*, troviamo talvolta degli arricciamenti sorprendenti. Questa, almeno, è la conclusione che si potrebbe trarre da uno degli esperimenti più ampiamente discussi – e criticati – nel campo della neuroscienza: gli esperimenti neurochirurgici di Benjamin Libet tesi a dimostrare ciò che egli chiama il «riferimento all'indietro nel tempo».

Talvolta durante gli interventi di chirurgia cerebrale è importante che il paziente sia sveglio e attento – si somministra quindi solo un anestetico locale (come la Novocaina che prendiamo dal dentista). Ciò permette al neurochirurgo di essere immediatamente informato dal paziente su ciò che sente mentre il suo cervello viene esaminato (vedi capitolo 3, nota 5). Questa tecnica è stata introdotta da Wilder Penfield (1958), e da più di trent'anni i neurochirurghi stanno raccogliendo dati sui risultati di una diretta stimolazione elettrica delle varie aree della corteccia. Si sa da molto tempo che la stimolazione di punti della corteccia *somatosensoria* (una fascia facilmente accessibile collocata nella parte superiore del cervello) produce nel paziente l'esperienza di sensazioni su parti corrispondenti del corpo. Per esempio, la stimolazione di un punto nell'area sinistra della corteccia somatosensoria può produrre la sensazione di un fugace pizzico nella mano destra del soggetto (per la ben nota inversione del sistema nervoso che lascia alla metà sinistra del cervello la responsabilità della parte destra del corpo, e viceversa). Libet ha confrontato l'andamento temporale di questi pizzichi indotti corticalmente con le sensazioni simili prodotte nel modo più usuale, applicando brevi impulsi elettrici alla mano stessa (Libet, 1965, 1981, 1982, 1985b; Libet e altri, 1979; vedi anche Popper ed Eccles, 1977; Dennett, 1979b; Churchland, 1981a, 1981b; Honderich, 1984).

Cosa ci aspetteremmo che accada? Supponiamo che per recarsi al lavoro due impiegati lascino casa alla stessa ora, ma uno vive in periferia e l'altro solo a pochi isolati dall'ufficio. Viaggiano alla stessa velocità, e quindi considerata la distanza supplementare che l'impiegato della periferia deve percorrere, arriverà in ufficio più tardi. Questo non è, comunque, quello che Libet ha scoperto quando ha chiesto ai suoi pazienti se è arrivato prima il pizzico alla mano partito dalla corteccia o quello partito normalmente dalla mano. Sulla base dei dati raccolti, egli ha affermato che sebbene passi sempre un tempo considerevole (approssimativamente 500 msec) tra l'inizio del-

6.2

la stimolazione e la «adeguatezza neuronale» (il punto in cui, Libet sostiene, i processi corticali culminano nella produzione dell'esperienza cosciente di un pizzico), quando la mano stessa è stimolata, l'esperienza viene «automaticamente riferita all'indietro nel tempo» e viene avvertita *prima* del pizzico prodotto dalla stimolazione cerebrale diretta.

Ancor più sorprendenti sono i casi riferiti da Libet in cui la *corteccia* sinistra del paziente veniva stimolata *prima* della sua *mano* sinistra, il che porterebbe a pensare che il soggetto dovesse sentire sicuramente due differenti pizzichi: prima alla mano destra (indotto corticalmente) e poi alla mano sinistra. In realtà, però, il resoconto soggettivo era rovesciato: «prima sinistra, poi destra».

Libet ha interpretato i suoi risultati come una seria sfida al materialismo: «[...] una dissociazione tra le determinazioni

temporali di eventi "mentali" e "fisici" corrispondenti sembrerebbe far sorgere delle difficoltà serie sebbene non insormontabili per la [...] teoria dell'identità psiconeurale» (Libet ed altri, 1979, p. 222). Secondo Sir John Eccles, premio Nobel per la medicina grazie alle sue ricerche in neurofisiologia, a questa sfida non si può far fronte:

Questa procedura di anticipazione non sembra essere spiegabile ricorrendo a un qualche processo neurofisiologico. Presumibilmente è una strategia appresa dalla mente auto-cosciente [...] l'anticipazione dell'esperienza sensibile è attribuibile all'abilità della mente autocosciente di compiere lievi aggiustamenti temporali, per esempio, truccare il tempo. [Popper e Eccles, 1977, p. 441]

Più recentemente il matematico e fisico Roger Penrose (1989) ha suggerito che una spiegazione materialistica dei fenomeni di Libet richiederebbe una rivoluzione dei fondamenti della fisica. Sebbene l'esperimento di Libet sia stato spesso salutato nei circoli non scientifici come una dimostrazione della verità del dualismo, poche persone nella comunità della scienza cognitiva condividono tale opinione. In primo luogo, le procedure sperimentali di Libet e le sue analisi dei risultati sono state severamente criticate; i suoi esperimenti non sono mai stati replicati, il che costituirebbe una ragione per molti aspetti sufficiente a privare di ogni considerazione i suoi «risultati». Il punto di vista scettico è, allora, che i suoi fenomeni semplicemente non esistono. *Ma se esistessero*? Questo è proprio il tipo di domanda che piace a un filosofo e comunque in questo caso, al di là dei gusti dei filosofi, c'è un altro motivo per porla. Nessuno dubita dell'esistenza di fenomeni più semplici come il *phi* colorato o il coniglio cutaneo, e la loro interpretazione fa sorgere esattamente gli stessi problemi. Sarebbe una miopia teoretica accettare le preoccupazioni *metodologiche* per rifiutare i fenomeni di Libet, dal momento che si lascerebbero intatte le presupposizioni di fondo che ci portano a pensare che *se* quell'esperimento *fosse* appropriatamente confermato, il materialismo subirebbe un duro colpo.

La prima cosa da notare sull'esperimento di Libet è che non presenterebbe nessuna anomalia se scartassimo la possibilità di registrare i resoconti verbali che i soggetti forniscono della loro esperienza, e quindi usarli per generare prima un testo e poi un mondo eterofenomenologico. I rumori che essi producono con i loro apparati vocali durante o dopo l'esperimento non fanno

venire in mente nessun paradosso se trattati come semplici fenomeni acustici. In nessun caso i suoni sembrano uscire dalle teste prima che le labbra si muovano, né le mani si muovono prima che si verifichino gli eventi cerebrali che chiaramente le hanno mosse, né gli eventi avvengono nella corteccia prima degli stimoli che si pensa li abbiano causati. Visti strettamente come il comportamento interno ed esterno di un sistema di controllo per il corpo biologicamente realizzato, gli eventi osservati e cronometrati negli esperimenti non mostrano alcuna violazione della ordinaria causalità meccanica – così come è approssimativamente descritta nel modello standard della fisica newtoniana e galileiana. Si potrebbe, allora, «far scomparire il problema» dichiarandosi comportamentisti estremi e rifiutando semplicemente di prendere sul serio i resoconti introspettivi.

Ma non siamo comportamentisti estremi; vogliamo accettare la sfida di dare un senso a ciò che Libet chiama «un aspetto fenomenologico primario della nostra esistenza umana in relazione alla funzione cerebrale» (1985a, p. 534). Libet si avvicina al nocciolo dell'eterofenomenologia, quando dice: «È importante rendersi conto che questi riferimenti e aggiustamenti soggettivi si svolgono evidentemente al livello della "sfera" *mentale*; non si evidenziano, come tali, nelle attività ai livelli neurali» (1982, p. 241). Ma poiché non ha a disposizione *alcun modo neutrale* per riferirsi alla fenomenologia, deve assegnare l'anomalia «alla "sfera" mentale». Un piccolo passo, un passo obbligato (dal momento che *deve* farlo se vuole rifiutare il comportamentismo), ma è il primo sulla sdrucciolevole china verso il dualismo.

> I resoconti dei soggetti sulle loro esperienze differenti [...] non erano costrutti teoretici ma osservazioni empiriche. [...] Il metodo dell'introspezione può avere i suoi limiti, ma può essere usato appropriatamente nell'ambito della scienza naturale ed è assolutamente essenziale se si vuole ottenere qualche dato sperimentale sul problema mente-corpo. (1987, p. 785)

I resoconti dei soggetti, anche quando sono trasformati in testi, sono, come dice Libet, delle osservazioni empiriche, ma *ciò che riferiscono*, gli eventi nei loro mondi eterofenomenologici, sono in realtà dei costrutti teoretici. Possono, come propone Libet, essere usati appropriatamente nell'ambito della scienza naturale, ma *solo* se si intendono fin dall'inizio come finzioni teoriche.

Libet sostiene che i suoi esperimenti di stimolazione diretta della corteccia dimostrano «due fattori temporali notevoli»:

1. *C'è un cospicuo ritardo prima che le attività cerebrali*, avviate dallo stimolo sensibile, *raggiungano «l'adeguatezza neuronale»* per la produzione di qualunque risultante esperienza sensibile cosciente.

2. Dopo che l'adeguatezza neuronale è raggiunta, *la determinazione soggettiva del tempo in cui l'esperienza è avvenuta viene (automaticamente) riferita indietro nel tempo*, utilizzando come «segnale temporizzatore» la risposta iniziale della corteccia cerebrale allo stimolo sensibile (1981, p. 182).

Il «segnale temporizzatore» è la *prima* esplosione di attività che appare nella corteccia (il primo potenziale evocato), che si verifica solo 10-20 msec dopo la stimolazione periferica dell'organo sensibile. Libet suggerisce che il riferimento all'indietro avviene sempre «al» segnale temporizzatore.

Il modello di Libet è staliniano: dopo che il primo potenziale è stato evocato, vari processi editoriali avvengono nella corteccia prima della «adeguatezza neuronale», e a quel punto un film completo viene proiettato. Come viene proiettato? Qui la descrizione di Libet oscilla tra una posizione estrema e una moderata (cfr. Honderich, 1984):

1. *Proiezione all'indietro*: il film è in qualche modo inviato indietro nel tempo in un Teatro Cartesiano dove viene proiettato in sincrono con i primi potenziali evocati. (I primi potenziali evocati, in quanto «segnali temporizzatori», svolgono la stessa funzione del ciak nella cinematografia, indicano esattamente al proiettore di quanto all'indietro nel tempo si deve inviare l'esperienza.)

2. *Riferimento all'indietro*: il film è proiettato nel tempo ordinario, ma presenta una specie di timbro, che ricorda allo spettatore di immaginare che quegli eventi siano avvenuti in un tempo precedente. (In questo caso i primi potenziali evocati servono semplicemente come date, che possono essere *rappresentate* sullo schermo cartesiano da didascalie come «Alla vigilia della battaglia di Waterloo» o «New York City, estate del 1942».)

Il termine usato da Libet è *riferimento*, ed egli lo difende ricordandoci il fenomeno «da molto tempo conosciuto e accettato» del riferimento spaziale, il che suggerisce una lettura moderata.

Il riferimento soggettivo all'indietro nel tempo è uno strano concetto e forse non sembra facilmente accettabile dapprincipio. Ma può vantare un precedente importante nel concetto da molto tempo conosciuto e accettato del riferimento soggettivo nella dimensione spaziale. Per esempio, l'immagine visiva di cui si ha esperienza a seguito di uno stimolo visivo ha una collocazione e una configurazione spaziale soggettiva che sono molto differenti dalla collocazione e configurazione spaziale delle attività neuronali che fanno sorgere l'immagine («riferita soggettivamente»). (1981, p. 183. Vedi anche Libet ed altri, 1979, p. 221; Libet, 1985b.)

Comunque, alla fine conclude dicendo che il riferimento temporale solleva dei problemi per il materialismo (per la «teoria dell'identità psiconeurale»: Libet ed altri, 1979, p. 222), cosicché o pensa che anche il riferimento spaziale solleva questi problemi o non ha capito la sua propria difesa. Se il riferimento spaziale – il fatto che ciò che vediamo sembra stare fuori, e non dentro, i nostri cervelli – solleva un problema per il materialismo, allora perché Libet sostiene che il suo lavoro rivela un argomento importante e nuovo a favore del dualismo? Sicuramente il riferimento spaziale costituisce un fatto molto meglio attestato di quella sorta di riferimento temporale che egli così ingegnosamente cerca di dimostare. Sembra, però, che Libet abbia una visione radicale (o in ogni caso confusa) del riferimento spaziale come di un certo tipo di «proiezione»:

esistono prove sperimentali a favore dell'ipotesi che la «sfera» mentale o soggettiva possa effettivamente «riempire» le lacune spaziali e temporali. In che modo si potrebbe altrimenti comprendere, per esempio, la già menzionata discrepanza enorme *che sappiamo sussistere* tra l'immagine visiva soggettiva e la configurazione di attività neuronali che fa sorgere l'esperienza dell'immagine? (1981, p. 196)

Ciò sembrerebbe equivalente a dire che il proiettore che Smythies non riusciva a trovare nel cervello si nasconde in realtà nella «sfera» mentale.[8]

Come fa Libet a stabilire i suoi due notevoli fattori temporali? «L'adeguatezza neuronale», che secondo Libet richiede fino a 500 msec di attività corticale, è determinata osservando

quanto tempo dopo una stimolazione iniziale *si possa ancora intervenire per influenzare* l'esperienza cosciente successivamente riferita. Oltre quell'intervallo critico, uno stimolo corticale diretto sarà riferito dal soggetto come un'esperienza *successiva*. (Essendo arrivato troppo tardi nell'ufficio editoriale per essere inserito nella «stampa finale» dell'esperienza del primo stimolo, apparirà nel prossimo numero.) I dati di Libet suggeriscono una finestra editoriale molto variabile: «Lo stimolo corticale poteva partire più di 500 msec dopo l'impulso cutaneo e modificare ancora la sensazione cutanea, benché nella maggior parte dei casi non sono stati osservati effetti retroattivi con intervalli (tra impulso periferico e impulso corticale condizionante) maggiori di 200 msec» (1981, p. 185). Libet è molto attento nel definire l'adeguatezza neuronale in termini di effetto sul successivo resoconto verbale non affrettato: «Al soggetto veniva chiesto di riferire, entro pochi secondi dall'applicazione di ciascuna coppia di [...] stimoli» (1979, p. 195), e sottolinea che «la determinazione temporale di un'esperienza soggettiva deve essere distinta da quella di una risposta comportamentale (come quella che avviene nel tempo di reazione) che potrebbe essere compiuta prima che la consapevolezza cosciente si sviluppi...» (Libet ed altri, 1979, p. 193).

Questa puntualizzazione gli permette di difendere un'interpretazione alternativa dei dati raccolti da Patricia Churchland. Churchland rappresenta il primo esempio di «neurofilosofo» (vedi il suo libro del 1986, *Neurophilosophy: Toward a Unified Science of the Mind-Brain*). Quando lessi per la prima volta i risultati di Libet (in Popper ed Eccles, 1977), la incoraggiai a studiarli, e lei li ha passati attentamente al setaccio (Churchland, 1981a). Ha tentato di screditare la prima tesi di Libet, il lungo tempo necessario a raggiungere «l'adeguatezza neuronale» per la coscienza, chiedendo ai soggetti in un esperimento di dire «via» non appena erano coscienti di uno stimolo cutaneo come quello usato da Libet. Ella riporta una risposta temporale media su nove soggetti di 358 msec, il che secondo lei dimostra che i soggetti dovevano aver raggiunto l'adeguatezza neuronale al più tardi 200 msec dopo lo stimolo (considerando il tempo necessario alla produzione di una risposta verbale).

La replica di Libet è staliniana: una reazione verbale – il dire «via» – può iniziare inconsciamente. «Non c'è nulla di magico o di particolarmente informativo nel fatto che una risposta motoria sia costituita da una vocalizzazione della parola "via" invece della più usuale pressione con il dito di un

pulsante [...] L'abilità di rilevare uno stimolo e di reagire appropriatamente ad esso, o di essere psicologicamente influenzato da esso, senza alcuna riferibile consapevolezza cosciente dello stimolo, è ampiamente accettata» (1981, pp. 187-188). E all'obiezione «ma che cosa stavano pensando di fare i soggetti di Churchland, se non dire, come richiesto, quando erano coscienti dello stimolo?», Libet potrebbe dare la risposta standard degli staliniani: essi sono effettivamente diventati coscienti dello stimolo alla fine, ma a quel punto il loro resoconto verbale era già stato avviato.[9]

Per questa ragione Libet rifiuta studi sul tempo di reazione come quelli di Churchland, ritenendo che abbiano «una validità dubbia come criterio primario di un'esperienza soggettiva» (1981, p. 188). Egli preferisce lasciare che il soggetto si prenda il suo tempo: «Il resoconto è formulato in modo non affrettato entro pochi secondi dopo ogni prova, permettendo al soggetto di esaminare introspettivamente i suoi dati» (p. 188). Come fa, allora, a contrastare la prospettiva antagonistica che questo passo flemmatico conceda al revisionista orwelliano nel cervello tutto il tempo che gli occorre per rimpiazzare le *reali* memorie della coscienza con *false* memorie?

> Per utilizzare i resoconti dopo le prove è ovviamente necessario che i processi di memoria a breve termine e di rammemorazione siano operativi, ma ciò non presenta difficoltà con soggetti senza difetti significativi in queste abilità. (p. 188)

Questa è una petizione di principio ai danni dell'orwelliano, che è pronto a spiegare svariati effetti come il prodotto di un *normale* malfunzionamento nella memoria o di un ricordo allucinatorio, nel quale un evento precedente e reale nella coscienza viene cancellato e rimpiazzato da memorie successive. È Libet che ha lasciato cuocere lo stufato troppo a lungo o è Churchland che lo ha assaggiato troppo presto? Se Libet desidera riservare uno status *privilegiato* alla sua scelta del tempo del sondaggio, deve essere pronto a combattere i controargomenti.

Libet si avvicina ad una dichiarazione di *non belligeranza*: «Si deve ammettere che un resoconto di un ordine temporale relativo non può, di per sé, fornire un indicatore del tempo "assoluto" (il tempo dell'orologio) dell'esperienza: come suggerito, non esiste un metodo conosciuto per raggiungere un tale indicatore» (1981, p. 188). Questo fa eco alla sua precedente osservazione che non sembrava esserci «nessun metodo con il

quale determinare il tempo assoluto in cui si verifica un'esperienza soggettiva» (Libet ed altri, 1979, p. 193). Ciò che Libet non capisce, però, è che forse questo avviene perché non esiste un simile momento del tempo assoluto (cfr. Harnad, 1989).

Anche Churchland, nelle sue critiche (1981a, 1981b) cade nell'errore di non distinguere il tempo rappresentato dal tempo del rappresentante:

> Le due ipotesi si differenziano essenzialmente solo sulla determinazione del momento in cui le rispettive sensazioni *vengono sentite.* [Il corsivo è mio; 1981a, p. 177]

> Anche se si suppone che le sensazioni che derivano dalle simultanee sensazioni della pelle e del LM [lemnisco mediale] vengano *sentite esattamente nello stesso tempo* [il corsivo è mio], il ritardo nell'adeguatezza neuronale per lo stimolo cutaneo può essere facilmente un artefatto della situazione. (1981b, p. 494)

Supponiamo che ogni artefatto del genere fosse eliminato, ma le sensazioni fossero *ancora* «sentite esattamente nello stesso tempo». Come interpreterebbe Churchland questo risultato inaspettato? Ciò significherebbe che esiste un momento *t* tale che lo stimolo 1 è sentito in *t* e lo stimolo 2 è sentito in *t* (la prospettiva antimaterialista) o solo che lo stimolo 1 e lo stimolo 2 sono sentiti (esperiti) come simultanei? Churchland non si oppone all'implicazione che le scoperte di Libet, se giustificate, porterebbero lo scompiglio (come egli talvolta pretende) nel campo del materialismo. Altrove, comunque, ella nota correttamente che «per quanto possano essere intriganti, non c'è ragione di supporre che nelle illusioni temporali ci sia qualcosa di preternaturale, e certamente non c'è nulla in loro che le distingua dalle illusioni spaziali o dalle illusioni del movimento, nessun segno che riveli una loro particolare origine non fisica» (1981a, p. 179). Questo sarebbe giusto solo se le illusioni temporali fossero fenomeni in cui il tempo *è rappresentato in modo errato*; ma se tali *errate rappresentazioni* avvengono in tempi «errati» è in gioco qualcosa di più rivoluzionario.

Cosa si può dire in definitiva degli esperimenti di Libet sulla stimolazione corticale? Costituiscono un tentativo interessante ma inconcludente per stabilire qualcosa sul *modo in cui il cervello rappresenta l'ordine temporale.* I primi potenziali evocati possono servire in qualche modo come punti di riferimento specifici per le rappresentazioni neurali del tempo, benché Libet non sia

riuscito a dimostrarlo, come le critiche tecniche di Churchland mostrano chiaramente. In alternativa, il cervello può trattare le proprie rappresentazioni temporali in modo più elastico. Noi non rappresentiamo gli oggetti visti come se esistessero sulla retina, ma piuttosto a varie distanze nel mondo esterno; perché mai il cervello non potrebbe rappresentare gli eventi come se avvenissero *nel momento in cui* ha più senso «ecologico» che avvengono? Quando siamo impegnati in qualche attività che richiede destrezza manuale, «il tempo sulla punta delle dita» dovrebbe essere lo standard; quando stiamo dirigendo un'orchestra, «il tempo dell'orecchio» potrebbe dominare l'ordinamento seriale. «Il tempo corticale di base» potrebbe essere quello standard (come il tempo di Greenwich per l'impero britannico), un problema, comunque, su cui si deve ulteriormente indagare.

L'argomento è stato oscurato dal fatto che entrambi gli antagonisti non hanno distinto chiaramente tra tempo del rappresentante e tempo rappresentato. È un dialogo tra sordi, con Libet che adotta la posizione staliniana e Churchland che fa le contromosse orwelliane, e tutti e due che sembrano d'accordo sul fatto che c'è un momento preciso in cui (un tempo «assoluto», come direbbe Libet) un'esperienza cosciente avviene.[10]

4. *Le tesi di Libet sulla dilazione soggettiva della coscienza dell'intenzione*

L'idea che esista una determinazione temporale assoluta di un'esperienza viene sfruttata anche negli esperimenti successivi di Libet sulle «intenzioni coscienti». In questi esperimenti egli ha cercato di determinare questo tempo assoluto lasciando che i soggetti stessi, gli unici ad avere (in qualche modo) un accesso diretto alle loro esperienze, eseguano un'*auto-temporizzazione*. Libet chiese a dei soggetti normali (non pazienti neurochirurgici) di prendere la decisione «spontanea» di flettere la mano all'altezza del polso notando al contempo la posizione di un punto su di un disco ruotante (la lancetta di un orologio, in pratica) nell'istante preciso in cui formulavano la decisione (Libet, 1985a, 1987, 1989). Successivamente (alcuni secondi più tardi), i soggetti riferivano in quale posizione si trovava il punto nel momento in cui avevano deciso di flettere il polso. Ciò permetteva a Libet di calcolare l'istante esatto (fino al millisecondo) in cui i soggetti *pensavano* di aver deciso e di confrontare

quel momento con lo svolgimento temporale degli eventi che contemporaneamente si verificavano nei loro cervelli. Egli scoprì che queste «decisioni coscienti» seguivano di 350-400 msec l'insorgere dei «potenziali di prontezza» che era in grado di registrare dagli elettrodi sul cranio, che secondo lui intercettavano gli eventi neurali determinanti le azioni volontarie eseguite. Ciò lo ha portato a concludere che «la preparazione cerebrale di un atto volontario spontaneo inizia inconsciamente» (1985a, p. 529).

Sembrerebbe quindi che la nostra coscienza si affanni dietro i processi cerebrali che effettivamente controllano il nostro corpo. Per molti questa è una prospettiva sconvolgente e perfino deprimente, giacché sembra escludere un vero (non solo illusorio) «ruolo esecutivo» per «l'io cosciente». (Vedi le discussioni dei molti commentatori di Libet, 1985a, 1987, 1989; e in Harnad, 1982; Pagels, 1988, pp. 233 sgg.; e Calvin, 1989a, pp. 80-81.)

Libet è più chiaro della maggior parte dei suoi critici sull'importanza di tener separati il contenuto e il veicolo: «Non si dovrebbe confondere *quello* che è riferito dal soggetto con *quando* egli può essere diventato introspettivamente consapevole di ciò che sta riferendo» (Libet, 1985a, p. 559). Egli riconosce (p. 560), inoltre, che un giudizio di simultaneità non deve esso stesso arrivare o essere reso simultaneamente; può maturare in un periodo più ampio di tempo (si consideri, per esempio, i minuti che occorrono ai giudici dell'ippodromo per sviluppare e poi esaminare il fotofinish sul quale basare il loro verdetto di vittoria o di parità).

Libet ha raccolto i dati di due serie temporali:

- la serie oggettiva, che include le determinazioni temporali dell'orologio esterno e degli eventi neurali salienti: i potenziali di prontezza (PP) e l'elettromiogramma (EMG) che registra l'avvio della contrazione muscolare;

- la serie soggettiva (così come è riferita successivamente), che consiste delle immagini mentali, delle memorie di qualsiasi pianificazione e di un unico punto di riferimento per ogni prova: un giudizio di simultaneità della forma *la mia intenzione cosciente (I) è iniziata quando la lancetta dell'orologio era nella posizione P.*

Sembra che Libet voglia tallonare lo sfuggente *acte gratuit*

discusso dagli esistenzialisti (per es., Gide, 1948; Sartre, 1943), la scelta completamente immotivata – e quindi in un certo qual senso «libera». Come numerosi commentatori hanno sottolineato, è molto difficile considerare tali azioni fortemente inusuali (che potrebbero essere chiamate atti di deliberata pseudo-casualità) come paradigmi dei «normali atti volontari» (Libet, 1987, p. 784). Ma, in ogni caso, è possibile ritenere che Libet abbia isolato una varietà di esperienza cosciente, comunque la si voglia caratterizzare, della quale si possa dare una temporalizzazione assoluta tramite un simile progetto sperimentale?

Egli sostiene che quando le intenzioni coscienti di agire (almeno quelle del suo tipo speciale) vengono messe in correlazione con gli eventi cerebrali che effettivamente danno l'avvio a quelle azioni, ci troviamo di fronte ad uno sfasamento di circa 300-500 msec. Questo ritardo è enorme – fino a mezzo secondo – e appare sinistro a chiunque sia fedele al principio che le nostre azioni coscienti *controllano* i nostri movimenti corporei. È come se *noi* fossimo all'interno del Teatro Cartesiano in cui ci viene mostrato, con un mezzo secondo di ritardo, il *reale* processo decisionale che si sta svolgendo *altrove* (in qualche luogo dove *noi* non siamo). Non siamo del tutto «fuori del giro» (come dicono alla Casa Bianca), ma poiché il nostro accesso alle informazioni è differito, il massimo che possiamo fare è intervenire con qualche «veto» dell'ultimo momento. Nel flusso che sgorga dal Quartier Generale Supremo, io non svolgo nessuna reale iniziativa, non mi trovo mai all'inizio di un progetto, ma esercito un minimo di influenza sulle politiche già formulate che passano nel mio ufficio.

Questa immagine è attraente ma incoerente. Il modello di Libet, come al solito, è staliniano, ma c'è un'ovvia alternativa orwelliana: i soggetti *erano* coscienti delle loro intenzioni fin dal primo momento, ma questa consapevolezza è stata spazzata via dalla memoria (o solo revisionata) prima che avessero la possibilità di ricordarla. Libet concede che ciò «presenta effettivamente un problema, ma non era controllabile sperimentalmente» (1985a, p. 560).

Una volta fatta questa concessione, non è mal concepito il tentativo di stabilire il tempo assoluto della coscienza? Né Libet né i suoi critici tirano questa conclusione. Libet, pur avendo distinto attentamente il contenuto dal veicolo – *quello* che viene rappresentato da *quando* viene rappresentato – cerca tuttavia di arrivare ad una conclusione sul tempo assoluto in cui il rappresentante si è presentato nella coscienza utilizzando come pre-

messa ciò che viene rappresentato. Lo psicologo Gerald Wasserman (1985, p. 556) individua il problema: «Il momento in cui il punto oggettivo esterno occupa una data posizione dell'orologio può essere determinato facilmente, ma questo non è il risultato desiderato». Ma successivamente cade nella trappola cartesiana: «Ciò di cui si ha bisogno è il momento in cui si verifica la rappresentazione mente-cervello interna del punto».

«*Il* momento in cui si verifica» la rappresentazione interna? Si verifica dove? C'è una rappresentazione essenzialmente continua del punto (che lo rappresenta in varie posizioni differenti) in varie parti differenti del cervello, partendo dalla retina e continuando attraverso il sistema visivo. La luminosità del punto viene rappresentata in certi luoghi e tempi, la sua posizione in altri, il suo movimento in altri ancora. Man mano che il punto esterno si muove, tutte queste rappresentazioni cambiano in modo asincrono e spazialmente distribuito. Dove «tutto ciò confluisce in un istante nella coscienza»? Da nessuna parte.

Wasserman osserva correttamente che il compito del soggetto di determinare dove il punto si trovasse in qualche momento della sequenza soggettiva è esso stesso un compito volontario, e per avviarlo occorre presumibilmente del tempo. Ciò crea difficoltà non solo perché è in competizione con altri progetti contemporanei, ma anche perché è innaturale – un tipo di giudizio cosciente sul tempo che normalmente non gioca un ruolo nel controllo del comportamento e quindi non ha un significato naturale nella sequenza. Il processo di interpretazione che alla fine fissa il giudizio di simultaneità soggettiva è esso stesso un artefatto della situazione sperimentale, e *muta il compito*, non dicendoci perciò nulla di interessante sulle determinazioni temporali reali dei veicoli della rappresentazione normale in qualunque luogo del cervello.

L'immagine fin troppo naturale che dobbiamo scartare è la seguente. In qualche luogo oscuro del cervello i preparativi per un'azione iniziano; dapprincipio si tratta solo di un'intenzione inconscia, ma mentre si avvicina lentamente al Teatro acquista determinatezza e potere, poi, all'improvviso, nell'istante *t*, balza sulla scena, dove una processione di rappresentazioni visive del punto sta sfilando – dopo essersi lentamente rivestite di lucentezza e determinatezza spaziale nel loro tragitto dalla retina. Al pubblico o all'*io* è dato il compito di dire esattamente quale rappresentazione del punto era «sul palcoscenico» quando l'intenzione cosciente ha fatto il suo inchino. Una volta

identificata la rappresentazione, si può calcolare il momento in cui il «suo» punto è partito dalla retina, così come la distanza dal teatro e la velocità di trasmissione. In questa maniera possiamo determinare il momento esatto in cui l'intenzione cosciente si è presentata nel Teatro Cartesiano.

È incredibile quanto sia attraente quest'immagine. È così facile da visualizzare! Sembra così appropriata! Non è proprio quello che *deve* accadere quando due cose si verificano contemporaneamente nella coscienza? No. In realtà, *non può* essere ciò che accade quando due cose si verificano contemporaneamente nella coscienza, perché non esiste un luogo simile nel cervello. Qualcuno ha pensato che l'incoerenza di *questa* descrizione non obbliga ad abbandonare l'idea di una determinazione assoluta del tempo delle esperienze. Sembra esistere un modello alternativo dell'insorgere della coscienza che evita l'assurdità cartesiana di un centro del cervello ma permette la temporizzazione assoluta. La coscienza potrebbe essere non una questione di arrivo ad un punto, ma piuttosto di rappresentazione che supera una certa soglia di attivazione sull'intera corteccia o su una larga parte di essa. Secondo questo modello, un elemento del contenuto diventa cosciente in un momento *t*, non entrando in qualche sistema definito funzionalmente e anatomicamente localizzato, ma cambiando stato lì dove si trova: acquistando qualche proprietà o avendo l'intensità di una delle sue proprietà spinta al di sopra di una certa soglia.

L'idea che la coscienza sia un *modo di azione* del cervello piuttosto che un suo *sottosistema* è molto apprezzabile (vedi, per es., Kinsbourne, 1980; Neumann, 1990; Crick e Koch, 1990). Inoltre, si può presumere che tali transizioni modali possano essere cronometrate da osservatori esterni, fornendo, in linea di principio, un'unica e determinata sequenza di contenuti che raggiunge il modo speciale. Ma con ciò rimaniamo ancora all'interno del Teatro Cartesiano, se si sostiene che il tempo reale («assoluto») di tali transizioni modali sia ciò che definisce la sequenza soggettiva. L'immagine è leggermente differente, ma le implicazioni sono le stesse. Attribuire la proprietà speciale che genera la coscienza a un determinato istante è solo metà del problema; stabilire che la proprietà è stata attribuita a quel momento è l'altra, e benché gli osservatori scientifici con i loro strumenti possano essere in grado di farlo con un'accuratezza al millisecondo, come può farlo il cervello?

Noi esseri umani formuliamo dei giudizi di simultaneità o successione degli elementi della nostra esperienza (alcuni li

esprimiamo anche), quindi in qualche punto o punti dei nostri cervelli viene girato l'angolo dal tempo della rappresentazione alla rappresentazione del tempo, e dopo che tali discriminazioni sono state effettuate in qualche luogo e qualche tempo, le proprietà temporali delle rappresentazioni che incorporano tali giudizi non sono costitutive del loro contenuto. Le oggettive simultaneità o successioni degli eventi che si verificano nell'ampio campo della corteccia non hanno alcuna rilevanza funzionale *se non possono essere accuratamente rilevate da qualche meccanismo cerebrale*. Possiamo far emergere il punto cruciale con una domanda: cosa trasformerebbe *questa* successione nel flusso della coscienza? Non c'è nessuno all'interno del cervello che osserva lo spettacolo che si svolge sull'ampio schermo della corteccia, anche se esso è osservabile da osservatori *esterni*. Ciò che conta è il modo in cui questi contenuti vengono utilizzati da o incorporati nei processi dell'incessante controllo del comportamento, e ciò *deve* essere condizionato solo limitatamente dal tempo corticale. Ciò che conta, ancora una volta, non sono le proprietà temporali dei rappresentanti, ma le proprietà temporali *rappresentate*, qualcosa determinato dal modo in cui sono «ricevute» dai successivi processi nel cervello.

5. *Un premio: la giostra precognitiva di Grey Walter*

Essendoci affaticati sui casi difficili, ci meritiamo di soffermarci su qualcosa di strano ma di *relativamente* facile da capire – qualcosa che comunica chiaramente il messaggio di questo difficile capitolo. L'esperimento di Libet sulla auto-temporizzazione, abbiamo appena visto, creava un compito (un giudizio) difficile e artificiale, che spogliava i risultati dei significati sperati. Un notevole esperimento precedente del neurochirurgo britannico W. Grey Walter (1963) non presenta questo inconveniente. Grey Walter eseguì il suo esperimento con pazienti sulla cui corteccia motoria aveva impiantato degli elettrodi. Egli voleva verificare l'ipotesi che l'avvio di certe attività registrate fossero la scintilla delle azioni intenzionali. Ogni paziente era posto di fronte a delle immagini create sullo schermo da un proiettore di diapositive con caricatore circolare. Il paziente poteva far avanzare il caricatore a suo piacimento, premendo il pulsante di controllo. (Si noti la somiglianza con l'esperimento di Libet: anche questa era una decisione «libera», determinata solo da una crescita endogena della noia o dalla curiosità per la

prossima diapositiva o dalla distrazione o da qualunque altro motivo.) Il paziente non sapeva, però, che il pulsante di controllo era finto, non era affatto collegato al proiettore! Ciò che in realtà faceva avanzare il caricatore delle diapositive era il segnale amplificato proveniente dagli elettrodi impiantati nella corteccia motoria del paziente.

Si potrebbe pensare che i pazienti non notassero nulla di anormale, ma in realtà erano molto stupiti, perché sembrava come se il proiettore stesse anticipando le loro decisioni. Dicevano che proprio mentre «stavano per» spingere il pulsante, ma prima di averlo effettivamente deciso, il proiettore passava alla diapositiva successiva – e quindi si trovavano a spingere il pulsante con la preoccupazione di far avanzare il caricatore di due diapositive! L'effetto era molto marcato, stando alla descrizione di Grey Walter; malgrado ciò sembra che egli non abbia mai eseguito l'esperimento supplementare a cui viene subito da pensare: introdurre una dilazione variabile per stabilire quanto tempo bisogna aggiungere per poter eliminare l'effetto da «giostra precognitiva».

Una differenza importante tra la situazione sperimentale di Grey Walter e quella di Libet sta nel fatto che il giudizio sull'ordinamento temporale che provoca la sorpresa nell'esperimento di Grey Walter fa parte di un normale compito di monitoraggio comportamentale. Sotto questo aspetto è più simile ai giudizi di ordinamento temporale tramite i quali il nostro cervello si accorge di un movimento da sinistra a destra invece che da destra a sinistra, piuttosto che ai giudizi di ordinamento «deliberati, coscienti». In questo caso il cervello «si aspetta» un feedback visivo sulla felice esecuzione del suo progetto di far avanzare il caricatore, ma il feedback arriva prima del previsto, scatenando un allarme. Questo potrebbe mostrarci qualche aspetto importante sulle effettive determinazioni temporali dei veicoli del contenuto e dei connessi processi cerebrali, ma, contrariamente a quanto sembrerebbe dapprincipio, non potrebbe dirci alcunché sulle «determinazioni temporali assolute della decisione cosciente di cambiare diapositiva».

Supponiamo, per esempio, che un esperimento supplementare a quello di Grey Walter mostrasse che una dilazione di 300 msec (come suggerito da Libet) dovrebbe essere incorporata nella realizzazione dell'atto per eliminare il senso soggettivo del cambiamento precognitivo di diapositiva. Ciò che una tale dilazione mostrerebbe in realtà sarebbe che le aspettative provocate dalla decisione di cambiare diapositiva sono regolate in

modo da aspettarsi un feedback visivo 300 msec più tardi e reagiscono inviando un allarme in altre circostanze. (Analogo ad un messaggio di shock da parte del comandante in capo di Calcutta a Whitehall a seguito della battaglia di New Orleans.) Il fatto che l'allarme viene *alla fine* interpretato nella sequenza soggettiva come una percezione di eventi disordinati (cambiamento prima della pressione del pulsante) non mostra nulla sul *momento* reale in cui si è verificata la coscienza della decisione di spingere il pulsante. La sensazione che i soggetti riferivano di provare, quella di non aver il tempo per «bloccare» l'avviata pressione del pulsante quando «vedevano che la diapositiva stava già cambiando» è una interpretazione naturale del cervello per sistemare (alla fine) i vari contenuti resi disponibili in vari tempi e incorporarli in una narrazione. Questa sensazione era già lì nel primo momento della coscienza dell'intenzione (nel qual caso l'effetto richiede una lunga dilazione prima «che si alzi il sipario» ed è staliniano) o era una retrospettiva reinterpretazione di un *fait accompli* che altrimenti creerebbe confusione (nel qual caso è orwelliano)? Spero sia ormai chiaro che le presupposizioni di tale domanda la squalifichino.

6. *Aspetti incompiuti*

Si potrebbe ancora voler obiettare che tutti gli argomenti di questo capitolo non riescano a rovesciare quest'ovvia verità: la nostra esperienza degli eventi avviene nel medesimo ordine in cui noi abbiamo esperienza che gli eventi avvengono. Se qualcuno pensa «Uno, due, tre, quattro, cinque», il suo pensare «uno» si verifica prima di pensare «due» e così via. Questo esempio illustra una tesi che è vera in generale, e che in effetti sembra senza eccezioni fintanto che restringiamo la nostra attenzione ai fenomeni psicologici «ordinari», di durata macroscopica. Ma gli esperimenti che abbiamo considerato riguardano eventi compressi in cornici temporali inusualmente anguste, di poche centinaia di millesecondi. In queste dimensioni, le presupposizioni ordinarie crollano.

Ogni evento nel tuo cervello ha una definita collocazione spazio-temporale, ma chiedere «Esattamente quando *tu* sei diventato cosciente dello stimolo?» presuppone che qualcuno di questi eventi è, o equivale a, il tuo divenire cosciente dello stimolo. Equivale a chiedere «Esattamente quando l'Impero Britannico è stato informato della tregua nella Guerra del

1812?». A un certo punto tra il 24 dicembre 1814 e il 15 gennaio 1815 – questo è sicuro; ma è assolutamente insensato voler scegliere un giorno e un'ora precisa. Anche se possiamo stabilire con precisione il momento in cui i vari funzionari dell'Impero sono stati informati, nessuno di questi momenti può essere scelto come la data in cui l'Impero Britannico è stato informato. La firma dell'armistizio è stato un atto ufficiale e intenzionale dell'Impero, ma la partecipazione delle forze britanniche nella battaglia di New Orleans ne è stato un altro, eseguito nella convinzione che nessun armistizio fosse stato firmato. Si potrebbe sostenere il principio che l'arrivo della notizia a Whitehall o Buckingham Palace a Londra debba essere considerato il momento ufficiale in cui l'Impero sia stato informato, poiché quello era il «centro nervoso» dell'Impero. Cartesio pensava che la ghiandola pineale fosse un tale centro nervoso nel cervello, ma si sbagliava. Poiché la cognizione e il controllo – e quindi la coscienza – è distribuita in tutta l'area del cervello, non è possibile scegliere nessun momento preciso come quello in cui avviene l'evento cosciente.

In questo capitolo, ho cercato di minare le immaginarie «basi» di alcune cattive abitudini di pensiero e di rimpiazzarle con concezioni migliori, ma lungo il percorso ho dovuto lasciare molti aspetti incompiuti. Il più sconcertante è, credo, quello rappresentato dall'affermazione metaforica che il «sondaggio» fa «precipitare le narrazioni». La temporalizzazione dei sondaggi indagatori da parte degli sperimentatori, ho sostenuto, può avere un sostanziale effetto revisionistico sul sistema di rappresentazione utilizzato dal cervello. Ma tra quelli che possono dirigere questo sondaggio indagatorio sul soggetto c'è il soggetto stesso. Se ti interessa sapere quando sei diventato cosciente di qualcosa, il tuo auto-esame e la tua auto-ricerca stabiliscono dei termini per nuove finestre di controllo, alterando quindi le limitazioni sui processi implicati.

I risultati dei sondaggi ad opera di estranei sono generalmente degli atti linguistici di un tipo o dell'altro, ed *esprimono giudizi su* vari contenuti della coscienza, che tutti possono ascoltare e interpretare. I risultati degli auto-sondaggi rientrano *nella stessa categoria semantica* – non «presentazioni» (nel Teatro Cartesiano) ma *giudizi* su come sembra al soggetto, giudizi che il soggetto stesso può poi interpretare, ricordare e tenere presenti nelle sue azioni. In entrambi i casi, questi eventi stabiliscono le interpretazioni di ciò che il soggetto esperisce, e così fornisce dei punti fissi nella sequenza soggettiva. Ma,

secondo il modello delle Molteplici Versioni, non c'è più spazio per la questione ulteriore se *in aggiunta* a un tale giudizio, e alle precedenti discriminazioni sulle quali è basato, c'è stata una *presentazione* dei materiali grezzi sottoposti allo scrutinio di un Giudice Superiore, il pubblico del Teatro Cartesiano. Questa idea non è ancora facilmente comprensibile, né tantomeno accettabile. Le strade che conducono ad essa devono essere ulteriormente appianate.

7

L'EVOLUZIONE DELLA COSCIENZA

Ogni cosa è ciò che è perché lo è diventato.
D'ARCY THOMPSON (1917)

1. *Dentro la scatola nera della coscienza*

La teoria abbozzata nell'ultimo capitolo ci fa fare qualche passo sulla strada della comprensione del modo in cui la coscienza possa risiedere nel cervello umano, ma il suo contributo principale è stato di carattere negativo: ha scosso la forza dittatoriale dell'idea del Teatro Cartesiano. Abbiamo iniziato a rimpiazzarla con un'alternativa precisa, ma certamente non siamo andati molto lontani. Per compiere ulteriori progressi dobbiamo cambiare campo e abbordare le complessità della coscienza da una prospettiva differente: quella dell'evoluzione. Poiché non è sempre esistita la coscienza umana, deve essere sorta da fenomeni precedenti che non presentavano i caratteri della coscienza. Forse se analizziamo ciò che deve – o può – essere stato implicato in quella transizione, capiremo meglio tali complessità e il loro ruolo nella creazione del fenomeno completamente sviluppato.

Nel suo elegante libretto, *I veicoli pensanti* (1984), il neuroscienziato Valentino Braitenberg descrive una serie di meccanismi autonomi sempre più complicati, costruendoli gradualmente a partire da congegni completamente senza vita e comicamente semplici fino ad arrivare a entità (immaginarie) che impressionano per il loro aspetto profondamente biologico e psicologico. Questo esercizio di immaginazione funziona grazie a quella che egli chiama la «legge della maggior fatica analitica»: è molto più facile immaginare il comportamento (e le implicazioni comportamentali) di un congegno che noi otteniamo grazie alla sintesi di meccanismi sempre più complessi, «andando dall'interno all'esterno», si potrebbe dire, piuttosto che provare ad analizzare il comportamento esterno di una «scatola nera» e immaginare cosa sta accadendo lì dentro.

Finora in effetti abbiamo trattato la coscienza come una

specie di scatola nera. Abbiamo preso per «dato» il suo «comportamento» (= fenomenologia) e ci siamo chiesti quale sorta di meccanismo nascosto nel cervello potesse spiegarlo. Rovesciamo ora la strategia e riflettiamo sull'evoluzione dei meccanismi cerebrali necessaria a fare questo e quello e vediamo se emerge qualcosa che ci dà un meccanismo plausibile per spiegare alcuni dei «comportamenti» enigmatici dei nostri cervelli coscienti.

Ci sono state molte teorie – o meglio, congetture – sull'evoluzione della coscienza umana, a partire dalle supposizioni di Darwin nell'*Origine dell'uomo* (1871). A differenza della maggior parte delle spiegazioni scientifiche, le spiegazioni evoluzionistiche sono essenzialmente delle narrazioni, che ci portano da un tempo in cui qualcosa non esiste a un tempo in cui esiste tramite una serie di passi che la narrazione ci spiega. Invece di cercare di esporre in modo accademico tutte le narrazioni che sono state escogitate, mi ripropongo di narrare una singola storia, attingendo liberamente da molti altri teorici, e concentrandomi su alcuni punti spesso trascurati che ci aiuteranno a superare gli ostacoli nella comprensione di cosa sia la coscienza. Per riuscire a raccontare una buona storia, e a mantenerla relativamente corta, ho resistito alla tentazione di includere dozzine e dozzine di affascinanti trame secondarie e ho rinunciato all'abituale istinto filosofico di saggiare i pro e i contro di ogni punto che accetto o rifiuto. Il risultato, lo riconosco, è un po' come un riassunto in cento parole di *Guerra e Pace*, ma ci aspetta ancora molto lavoro.[1]

La storia che dobbiamo raccontare è analoga ad altre storie che la biologia sta cominciando a narrare. Si pensi, per esempio, alla storia dell'origine del sesso. Esistono oggi molti organismi che non hanno genere e si riproducono asessualmente, e c'era un tempo in cui tutti gli organismi esistenti mancavano di genere, maschile e femminile. In qualche modo, grazie a una serie immaginabile di passi, alcuni di questi organismi devono essersi evoluti in organismi dotati di genere, ed infine, naturalmente, in noi. Quale tipo di condizioni erano richieste per favorire o rendere necessarie queste innovazioni? Perché, in breve, tutti questi cambiamenti sono avvenuti? Questi sono alcuni dei più profondi problemi nella teoria contemporanea dell'evoluzione.[2]

C'è un parallelismo interessante tra le due questioni, quella delle origini del sesso e quella delle origini della coscienza. Non c'è quasi nulla di *erotico* (in termini umani) nella vita sessuale

dei fiori, delle ostriche e di altre semplici forme di vita, ma possiamo riconoscere nelle loro routine riproduttive, meccaniche e apparentemente poco gioiose, le basi e i principi del nostro molto più eccitante mondo sessuale. Similmente, non c'è nulla di particolarmente *egoico* (se così si può dire) nei primitivi precursori dei sé umani coscienti, ma essi gettano le basi per le nostre specifiche innovazioni e complicazioni umane. La struttura della nostra mente cosciente è il risultato di tre successivi processi evolutivi, innestati uno sull'altro, ognuno molto più veloce e potente del suo predecessore; per poter comprendere questa piramide di processi, dobbiamo cominciare dall'inizio.

2. *Gli albori*

Scena Prima: La nascita dei confini e delle ragioni

In principio, non c'erano ragioni; c'erano solo cause. Nulla aveva uno scopo, nulla aveva qualcosa che assomigliasse sia pur lontanamente ad una funzione; il mondo era del tutto privo di teleologia. È facile comprenderne il perché: non esisteva nulla che avesse interessi. Ma dopo svariati millenni emersero dei semplici *replicatori* (R. Dawkins, 1976; vedi anche Monod, 1972, cap. 1, «Strani Oggetti»). Benché *essi* non avessero la più pallida idea dei loro interessi, e forse propriamente parlando non avessero interessi, noi, scrutando quei giorni dal nostro vantaggioso e divino punto di osservazione, possiamo assegnare loro in modo non arbitrario determinati interessi – generati dal loro «interesse» per l'autoriproduzione (ricavabile dalla loro definizione). Cioè, se essi riuscissero o no a replicarsi forse non faceva nessuna differenza, non era importante, non interessava a nessuno o a nessuna cosa (benché sembra proprio che dovremmo essere grati del fatto che ci riuscirono), ma perlomeno possiamo assegnare loro degli interessi in modo condizionale. *Se* questi semplici replicatori devono sopravvivere e replicarsi, cioè conservarsi di fronte ad una crescente entropia, il loro ambiente deve offrire certe condizioni: condizioni favorevoli alla replicazione devono essere presenti o perlomeno frequenti.

Volendoci esprimere in termini più antropomorfici, se questi semplici replicatori vogliono continuare a replicarsi, devono sperare in, e tendere a, varie cose; devono evitare le cose «cattive» e perseguire quelle «buone». Quando arriva sulla

scena una entità capace di un comportamento che allontana, sia pur molto primitivamente, la propria decomposizione e dissoluzione, essa porta con sé nel mondo il proprio «bene». Crea, cioè, un punto di vista dal quale è possibile dividere gli eventi del mondo in favorevoli, sfavorevoli e indifferenti. E la propria inclinazione innata a cercare la prima, evitare la seconda e ignorare la terza contribuisce in modo essenziale alla definizione delle tre classi. Non appena una creatura arriva quindi ad avere degli interessi, il mondo e i suoi eventi iniziano a creare delle *ragioni* per essa – indipendentemente dal fatto che la creatura le possa riconoscere pienamente (Dennett, 1984a). Le prime ragioni precedono il loro riconoscimento. In effetti, il primo problema da risolvere per i primi risolutori di problemi era quello di imparare a riconoscere le ragioni che la loro stessa esistenza ha portato ad esistere, e quindi ad agire di conseguenza.

Non appena qualcuno si assume l'incombenza della propria conservazione, i confini cominciano ad essere importanti, giacché se ti accingi a conservarti, non vuoi sprecare le energie tentando di conservare il mondo intero: tracci una linea. Diventi, in una parola, *egoista.* Questa forma primordiale di egoismo (che essendo primordiale è priva degli aspetti principali del *nostro* tipo di egoismo) costituisce uno dei contrassegni della vita. È difficile stabilire dove un blocco di granito finisce e un altro comincia; la frattura che genera il confine può essere abbastanza reale, ma non c'è nulla a difesa del territorio, nessuno che fa indietreggiare la frontiera o si ritiri. «Io contro il mondo» – questa distinzione tra ogni cosa all'interno di un confine chiuso e ogni cosa nel mondo esterno – sta al cuore di tutti i processi biologici, non solo dell'ingestione e dell'escrezione, della respirazione e della traspirazione. Si consideri, per esempio, il sistema immunitario, con i suoi milioni di anticorpi differenti schierati in difesa del corpo contro milioni di differenti intrusi stranieri. Questo esercito deve risolvere il problema fondamentale del riconoscimento: distinguere il proprio io (e i suoi amici) da tutto il resto. Il problema è stato risolto in modo molto simile a quello in cui le nazioni umane, e i loro eserciti, hanno risolto il loro: attraverso delle procedure standardizzate e meccanizzate di identificazione – a livello macroscopico, i passaporti e gli uffici doganali; a livello microscopico, le forme molecolari e i rilevatori di forma. È importante notare che questo esercito di anticorpi non ha generali, né Quartier Generale con un piano di battaglia, e neppure una descrizione del

nemico: gli anticorpi si rappresentano il nemico solo nella stessa maniera con la quale un milione di serrature si rappresentano le chiavi che le aprono.

Dovremmo notare vari altri fatti che sono già chiari in questo primo stadio. Sebbene l'evoluzione dipenda dalla storia, Madre Natura non è snob, non si lascia impressionare dalle origini. Non importa come e da dove un organismo ha acquistato il suo valore; il valore conta di più del blasone. Per quanto ne sappiamo, naturalmente, i pedigree dei primi replicatori sono più o meno gli stessi: ognuno di loro è il prodotto di una serie cieca e fortuita di selezioni. Ma se un iper-ingegnere capace di viaggiare nel tempo inserisse un *robot replicatore* in quell'ambiente, e se il valore del robot fosse uguale o migliore di quello dei suoi concorrenti naturali, i suoi discendenti potremmo ritrovarceli fra noi – potremmo perfino *essere* noi! (Dennett, 1987a, 1990b).

La selezione naturale non può dire come un sistema è diventato quello che è diventato, ma ciò non significa che non potrebbero esserci delle profonde differenze tra i sistemi «progettati» dalla selezione naturale e quelli progettati da ingegneri intelligenti (Langton, Hogeweg, in Langton, 1989). Per esempio, i progettisti umani, essendo lungimiranti ma con i paraocchi, tendono ad essere frustrati nei loro progetti da imprevisti effetti collaterali e interazioni, così cercano di evitarli dando ad ogni elemento del sistema un'unica funzione e di isolarlo dagli altri elementi. In contrasto, Madre Natura (il processo di selezione naturale) è notoriamente miope e senza obiettivi. Poiché non prevede affatto, non ha il modo di preoccuparsi di effetti collaterali non previsti. Non «tentando» di evitarli, sperimenta progetti in cui molti di questi effetti collaterali si verificano; molti di questi progetti sono terribili (chiedi a qualunque ingegnere), ma ogni tanto si verifica un *effetto collaterale inaspettatamente felice*: due o più sistemi funzionali non correlati interagiscono per produrre un premio: molteplici funzioni per gli stessi elementi. Le funzioni molteplici non sono assenti dagli artefatti umani, ma sono relativamente rare; in natura sono ovunque, e, come vedremo, una delle ragioni per cui i teorici hanno avuto tante difficoltà a trovare plausibili schemi progettuali della coscienza nel cervello è dovuta al fatto che hanno avuto la tendenza a pensare agli elementi cerebrali come portatori di una sola funzione ognuno.[3]

Abbiamo così posto le prime pietre; siamo ora in grado di spiegare i seguenti fatti primordiali:

1. Ci sono ragioni da riconoscere.
2. Dove ci sono ragioni, ci sono punti di vista da cui riconoscerle e valutarle.
3. Ogni agente deve riconoscere un «qui dentro» da un «mondo esterno».
4. Tutti i riconoscimenti devono essere svolti in ultima analisi da una miriade di procedure «cieche, meccaniche».
5. Dentro il territorio difeso non deve esserci sempre un Dirigente Capo o un Quartier Generale.
6. In natura il valore conta più del blasone; l'origine non conta.
7. In natura gli elementi spesso giocano funzioni multiple all'interno dell'economia di un singolo organismo.

Abbiamo già visto degli echi di questi fatti primordiali, nella ricerca del definitivo «punto di vista dell'osservatore cosciente» e nelle varie circostanze in cui abbiamo sostituito gli omuncoli con (squadre di) semplici meccanismi. Ma, come abbiamo visto, anche se il punto di vista dell'osservatore cosciente non è identico ai punti di vista primordiali dei primi replicatori che divisero i loro mondi in bene e male; ne è comunque un sofisticato discendente. (Dopo tutto, anche le piante hanno punti di vista in questo senso primordiale.)

Scena Seconda: Nuove e migliori maniere per produrre futuro

> E una delle più profonde, delle più generali funzioni degli organismi viventi è quella di prevedere, di produrre futuro come diceva Paul Valéry.
>
> FRANÇOIS JACOB (1982), p. 66

> Predire il futuro di una curva significa eseguire una certa operazione sul suo passato. L'operazione effettivamente divinatrice non è effettuata da nessun apparato, ma si possono invece costruire apparecchi che realizzano operatori capaci di dare una certa approssimazione.
>
> NORBERT WIENER (1948)

Nell'ultimo capitolo ho accennato di sfuggita al fatto che lo scopo principale del cervello è quello di produrre futuro; quest'affermazione merita ora un'attenzione maggiore. Per fronteggiare con successo le difficoltà un organismo o deve corazzarsi (come un albero o una tellina) e «sperare per il meglio» o deve

sviluppare dei metodi per sfuggire al pericolo e rifugiarsi nell'ambiente sicuro più vicino. Se scegli quest'ultima possibilità, ti trovi di fronte al problema primario che ogni *agente* deve continuamente risolvere:

Ora, che faccio?

Per risolvere questo problema, ti serve un sistema nervoso, per *controllare* le tue attività nel tempo e nello spazio. La giovane ascidia vaga nel mare alla ricerca di uno scoglio o un corallo su cui attaccarsi ed eleggerlo a sua dimora permanente. Per questo compito, ha un rudimentale sistema nervoso. Quando trova il suo posto e mette radici, non ha più bisogno del cervello, e quindi se lo mangia! (È come vincere un concorso a cattedra.)[4] La chiave del controllo risiede nell'abilità di *rilevare* o perfino *anticipare* le caratteristiche importanti dell'ambiente, così tutti i cervelli sono, essenzialmente, delle *macchine anticipatrici*. La conchiglia della tellina è un'eccellente corazza, ma non può rimanere sempre chiusa; il riflesso rigidamente cablato che fa chiudere di colpo la conchiglia è un rudimentale ma efficace anticipatore/evitatore di danni.

Ancora più primitive sono le risposte di avvicinamento e arretramento degli organismi più semplici; essi sono collegati alle fonti del bene e del male nel modo più semplice che si possa immaginare: le *toccano*. Allora, a seconda se la cosa toccata sia buona o cattiva per loro, o la respingono o la inghiottono (solo al momento opportuno, se sono fortunati). Fanno questo semplicemente perché sono «cablati» in modo tale che l'effettivo contatto con le caratteristiche buone o cattive del mondo scatena l'appropriato spasmo reattivo. Questo fatto sta alla base, come vedremo, di alcune tra le caratteristiche più terribili e squisite (letteralmente) della coscienza. In principio, *tutti* i «segnali» scatenati dalle cose dell'ambiente significano o «svignatela!» o «buttatici!» (Humphrey, di prossima pubblicazione).

Nessun sistema nervoso, in questo stadio primitivo, può usare un «messaggio» più spassionato o obiettivo che semplicemente lo *informi*, in modo neutrale, di qualche condizione. Ma i sistemi nervosi di una tale semplicità non possono ottenere molti vantaggi nel mondo, essendo capaci soltanto di ciò che potremmo chiamare un'*anticipazione prossima*: un comportamento che è appropriato all'*immediato* futuro. Cervelli migliori sono quelli che possono estrarre più informazioni, più veloce-

mente, e usarle in primo luogo per *evitare* il contatto nocivo o per *cercare* qualcosa di nutriente (e le possibilità di accoppiamento, non appena compare il sesso).

Messi di fronte al compito di estrapolare un utile futuro dal nostro passato personale, noi organismi cerchiamo di ottenere qualcosa senza pagare (o per lo meno a buon mercato): trovare le leggi del mondo – e se non ci sono, trovare le leggi approssimative del mondo – qualunque cosa ci dia un vantaggio. Da un certo punto di vista appare del tutto singolare che noi organismi otteniamo qualcosa dalla natura. C'è qualche ragione profonda per cui la natura dovrebbe scoprire le sue carte o rilevare le proprie regolarità a un'indagine occasionale? Ogni buon produttore di futuro può essere considerato una specie di trucco – un sistema improvvisato che guarda caso, il più delle volte, funziona; un colpo fortunato rispetto ad una regolarità che può essere rilevata nel mondo. Ogni fortunato anticipatore di questo tipo in cui Madre Natura si imbatte è destinato ad essere ricompensato, ovviamente, se migliora le caratteristiche dell'organismo.

All'estremo «minimalista», allora, abbiamo le creature che si rappresentano il meno possibile: quanto basta a *permettere al mondo di avvertirli* quando stanno per iniziare a fare qualcosa di sbagliato. Le creature che seguono questa politica non fanno piani: si buttano a capofitto e se qualcosa comincia a danneggiarli, «ne hanno abbastanza» e si ritirano – ma questo è il massimo che sanno fare.

Il passo successivo implica l'anticipazione a medio termine – per esempio l'abilità di schivare un sasso. Questo tipo di talento anticipatore è spesso «già cablato» – fa parte di un meccanismo innato progettato nel corso dei millenni per rilevare il tipo di connessione regolare (con eccezioni) che possiamo notare tra *le cose che si avvicinano o si ingrandiscono all'improvviso* e *le cose che ci colpiscono*. La risposta che consiste nel ritrarsi repentinamente di fronte a qualcosa che si avvicina all'improvviso è cablata in modo rigido negli esseri umani, per esempio, e può essere osservata nei neonati (Yonas, 1981); costituisce un dono proveniente dai nostri remoti antenati che avevano dei cugini che non sopravvissero perché non ne sapevano abbastanza sullo schivare i sassi. Il segnale «qualcosa si avvicina improvvisamente» *significa* «abbassati!»? Be', questo potrebbe essere il suo proto-significato; è collegato direttamente al meccanismo che fa abbassare la testa.

Abbiamo altri doni del genere. I nostri sistemi visivi, come

quelli di molti altri animali, perfino i pesci, sono particolarmente sensibili alle configurazioni con assi di simmetria verticali. Braitenberg suggerisce che ciò è dovuto probabilmente al fatto che nel mondo naturale dei nostri remoti antenati (molto tempo prima che ci fossero le facciate delle chiese e i ponti sospesi), le uniche cose – o quasi – che *mostrassero* assi di simmetria verticali erano altri animali e solo quando stavano guardandoti. Così i nostri antenati sono stati dotati di sistemi di allarme estremamente preziosi che scattavano (quasi) ogniqualvolta erano *osservati* da un altro animale (Braitenberg, 1984).[5] Identificare un predatore a una certa distanza (spaziale) – invece di dover aspettare di sentire i suoi denti nel tuo collo – significa anche avere una specie di anticipazione più distale *in termini temporali*: ti dà un vantaggio nella corsa per la salvezza.

Una caratteristica importante di questi meccanismi è la loro rozzezza discriminativa; preferiscono rinunciare a ciò che può essere chiamata *la verità e l'accuratezza descrittiva* in cambio *della velocità e dell'economia*. Alcune cose che fanno scattare un rilevatore di simmetria verticale non hanno una reale rilevanza per l'organismo: il raro aspetto quasi simmetrico di un albero o di un cespuglio, oppure (in tempi moderni) molti artefatti umani. Quindi la classe delle cose individuate da un tale meccanismo è costituita da una ciurma variegata – in cui dominano gli «animali che guardano verso di me» ma si trova anche un numero indefinitamente alto di falsi allarmi (rispetto a *quel* messaggio). Inoltre non tutte o non solo le configurazioni con simmetria verticale metteranno in moto il meccanismo; alcune configurazioni simmetricamente verticali non riusciranno per un motivo o per l'altro a farlo scattare, e ci saranno anche dei falsi allarmi; è il prezzo da pagare per un meccanismo rapido, economico e portatile, un prezzo che gli organismi, nel loro narcisismo (Akins, 1989), pagano volentieri. Questo fatto è facile da comprendere, ma alcune delle sue implicazioni per la coscienza non sono dapprincipio così ovvie. (Nel capitolo 12 questo aspetto diventerà importante quando ci chiederemo, per esempio: quali proprietà rileviamo con la nostra visione del colore? Cosa hanno in comune le cose rosse? E anche, perché il mondo ci appare così com'è?)

Essere informato (in modo fallibile) che un altro animale ti sta guardando è quasi sempre un evento molto significativo nel mondo naturale. Se l'animale non ti vuole mangiare, potrebbe essere un possibile compagno per l'accoppiamento o un rivale per l'accoppiamento o una preda che si è accorta del tuo arrivo.

L'allarme dovrebbe poi mettere in moto gli analizzatori «amico, nemico o preda?», di modo che l'organismo possa distinguere tra messaggi come: «Un tuo simile ti sta osservando!», «Un predatore ti ha preso di mira!», «La tua cena sta per svignarsela!». In alcune specie (per esempio, in alcuni pesci) il rilevatore di simmetria verticale è congegnato in modo tale da provocare un'improvvisa interruzione momentanea dell'attività in corso; tale interruzione è conosciuta come *risposta di orientamento*.

Lo psicologo Odmar Neumann (1990) suggerisce che le risposte di orientamento costituiscano la controparte biologica dell'allarme marinaresco «Tutti gli uomini in coperta!». La maggior parte degli animali, come noi, svolgono attività che vengono controllate in modo meccanico, «in pilota automatico», non utilizzando le loro capacità a pieno, ma lasciando il controllo a sottosistemi specializzati dei loro cervelli. Quando scatta un allarme specializzato (come, per noi, quello di un oggetto che si avvicina all'improvviso o quello della simmetria verticale) o qualsiasi allarme generale provocato da qualcosa di sorprendente o solo inaspettato, il sistema nervoso dell'animale viene mobilitato per trattare la possibilità di un'emergenza. L'animale interrompe le sue attività e fa un rapido esame o aggiornamento che dà ad ogni organo sensoriale l'opportunità di contribuire all'insieme dell'informazione disponibile e pertinente. Una *temporanea* arena centralizzata del controllo viene instaurata tramite l'attività neurale potenziata – tutte le linee sono aperte, per un breve periodo. Se il risultato di questo aggiornamento è l'avvio di un «secondo allarme», l'intero corpo dell'animale viene mobilitato, tramite una scarica di adrenalina; altrimenti l'attività potenziata rapidamente recede, i marinai fuori servizio tornano a dormire e gli specialisti riprendono le loro funzioni di controllo. Questi brevi episodi di interruzione delle normali attività e di vigilanza potenziata non sono in se stessi episodi di «consapevolezza cosciente» simile a quella umana (come spesso si ritiene esageratamente) o almeno non sono necessariamente esempi di un tale stato, ma sono probabilmente i necessari precursori, nell'evoluzione, dei nostri stati coscienti.

Neumann avanza l'ipotesi che queste risposte di orientamento iniziarono come reazioni a segnali di allarme, ma si dimostrarono così utili nel provocare un aggiornamento generalizzato che gli animali iniziarono sempre più frequentemente ad entrare nella modalità di orientamento. I loro sistemi nervosi

avevano bisogno di una modalità «Tutti gli uomini in coperta!», ma una volta ottenuto, costava poco o nulla avviarlo più di frequente e ripagava abbondantemente con le informazioni accresciute sullo stato dell'ambiente o sullo stato interno dell'animale. Si potrebbe dire che divenne un'abitudine, non più solamente sotto il controllo degli stimoli esterni, ma con un'origine interna (quasi come regolari esercitazioni antincendio).

La regolare vigilanza si è trasformata gradualmente in *esplorazione* regolare, e una nuova strategia comportamentale ha iniziato ad evolversi: l'acquisizione di informazione «fine a se stessa», nel caso possa dimostrarsi utile un giorno. Molti mammiferi furono attratti da questa strategia, specialmente i primati, i quali svilupparono degli occhi dotati di grande mobilità, che, grazie ai movimenti saccadici, forniscono uno scandaglio quasi continuo del mondo. Questo fatto segnò una svolta fondamentale nell'economia degli organismi che fecero questo balzo: la nascita della curiosità o della fame epistemica. Invece di raccogliere informazione solo per usarla immediatamente, informazione «usa e getta», essi cominciarono a diventare, come si è espresso lo psicologo George Miller, *informivori*: organismi affamati di sempre maggiori informazioni sul mondo che abitano (e su se stessi). Ma per far questo non hanno inventato o messo in campo dei sistemi completamente nuovi per raccogliere informazioni; come al solito nell'evoluzione, hanno creato questi nuovi sistemi rabberciando insieme le attrezzature ricevute in eredità dai loro antenati. Ciò ha lasciato le sue tracce, soprattutto sulle implicazioni emotive o affettive più sottili della coscienza, giacché sebbene le creature più evolute siano diventate ora raccoglitrici «disinteressate» di informazioni, i loro «inviati speciali» sono semplicemente il ridispiegamento dei sistemi di allarme e di incitamento dei loro antenati, che non mandano mai alcun messaggio «direttamente», ma hanno sempre l'abitudine d'inserire qualche «commento editoriale» positivo o negativo su qualsiasi informazione forniscano. Togliendo le virgolette e le metafore: i legami innati tra gli stati informazionali e le reazioni ad arretrare e inghiottire, evitare e perseverare, non sono stati spezzati, ma solo attenuati e ridiretti. (Torneremo su questo punto nel capitolo 12.)

Nei mammiferi, questo sviluppo evoluzionistico è stato favorito da una divisione del lavoro nel cervello che ha creato due aree specializzate: (approssimativamente) la *dorsale* e la *ventrale*. (Ciò che segue è un'ipotesi del neuropsicologo Marcel Kinsbourne.) Al cervello dorsale erano affidate le responsabilità di

pilotaggio «in linea», le responsabilità di tenere il vascello (il corpo dell'organismo) fuori pericolo; come le procedure per «il rilevamento delle collisioni» inserite nei videogiochi, esso doveva quasi continuamente scandagliare le cose che si avvicinavano o si allontanavano e in generale si doveva occupare di impedire che l'organismo sbattesse contro qualcosa o cadesse da un dirupo. Questo permetteva al cervello ventrale di avere un po' di tempo libero per concentrarsi sull'identificazione di vari oggetti nel mondo; poteva permettersi di studiare da vicino i particolari e analizzarli con cura, in un modo lento e seriale, perché poteva essere sicuro che il sistema dorsale si occupava di tenere il vascello lontano dagli scogli. Nei primati, secondo la congettura di Kinsbourne, questa specializzazione dorsale/ventrale si è alterata ed è evoluta ulteriormente nella nota specializzazione tra emisfero destro e emisfero sinistro: l'emisfero destro, globale e spaziotemporale; l'emisfero sinistro, analitico, seriale e più concentrato.

Noi abbiamo indagato solo un filo nella storia evoluzionistica del sistema nervoso e ci siamo serviti del più basilare meccanismo evoluzionistico: la selezione di particolari *genotipi* (combinazioni di geni) che hanno dimostrato di produrre individui (*fenotipi*) più adatti dei genotipi alternativi. Gli organismi con la fortuna di essere meglio «cablati» alla nascita tendono a produrre una progenie che sopravvive meglio e più numerosamente, così un buon «cablaggio rigido» si diffonde nella popolazione. E noi abbiamo delineato una progressione nello spazio progettuale dai più semplici rilevatori immaginabili di bene e di male a organismi così strutturati nell'architettura da avere una considerevole capacità di produrre utili anticipazioni in un ambiente relativamente stabile e prevedibile.

Nella prossima fase della nostra storia dovremo introdurre una notevole innovazione: l'emergenza di fenotipi individuali il cui interno non è cablato in modo del tutto rigido, ma si dimostra piuttosto variabile o *plastico*, e che quindi possono imparare, durante la loro propria vita. L'emergenza della plasticità nei sistemi nervosi si è verificata (più o meno) nello stesso tempo dello sviluppo che abbiamo appena abbozzato, e ha fornito due nuovi mezzi in cui l'evoluzione può avere luogo, ad una velocità molto maggiore che nella spontanea evoluzione genetica tramite mutazione dei geni e selezione naturale. Poiché alcune delle complessità della coscienza umana sono il risultato degli sviluppi che si sono verificati, e continuano a verificarsi, in questi nuovi mezzi, abbiamo bisogno di un'immagine chiara,

seppur elementare, delle relazioni che hanno tra di loro e con il sottostante processo di evoluzione genetica.

3. *L'evoluzione nei cervelli e l'effetto Baldwin*

Noi tutti presupponiamo che il futuro sarà come il passato – questa è la premessa indimostrabile ma essenziale di tutte le nostre inferenze induttive, come Hume ha notato. Madre Natura (lo sviluppo progettuale realizzato nel processo della selezione naturale) fa la stessa presupposizione. Sotto molti aspetti, le cose permangono identiche: la gravità continua ad esercitare la sua forza, l'acqua continua ad evaporare, gli organismi continuano ad avere bisogno d'acqua per riempire e proteggere i loro corpi, le cose che si avvicinano rapidamente continuano a sottendere porzioni sempre più grandi di retina, e così via. Quando sono implicati fatti generali come questi, Madre Natura fornisce delle soluzioni a lungo termine ai problemi: cabla rigidamente dei rilevatori «sopra-sotto» basati sulla gravità, cabla rigidamente dei «campanelli d'allarme» per la sete, cabla rigidamente dei circuiti per «abbassarsi quando un oggetto si avvicina all'improvviso». Altre cose cambiano, ma in modo prevedibile, in cicli, e Madre Natura risponde con altri congegni cablati rigidamente, come i meccanismi per far crescere la pelliccia invernale messi in funzione dai cambiamenti della temperatura e le sveglie interne per regolare i cicli di sonno e di veglia degli animali diurni e notturni. Ma talvolta le opportunità e le vicissitudini dell'ambiente sono relativamente imprevedibili sia per Madre Natura che per chiunque – sono, o sono influenzati da, processi *caotici* (Dennett, 1984a, pp. 109 sgg.). In questi casi, nessun progetto stereotipato si adatterà a tutte le eventualità, così gli organismi migliori saranno quelli che possono in una certa misura *riprogettare se stessi* per fronteggiare le condizioni che incontrano. Talvolta tale riprogettazione viene chiamata *apprendimento* e talvolta viene chiamata solo *sviluppo*. La linea divisoria tra i due è controversa. Gli uccelli *imparano* a volare? Imparano a cantare le loro canzoni? (Nottebohm, 1984; Marler e Sherman, 1983). E a far crescere le loro piume? I bambini *imparano* a camminare o a parlare? Poiché la linea divisoria (se c'è) non è importante per i nostri scopi, limitiamoci a chiamare tutti questi processi, nello spettro che va dall'imparare a mettere a fuoco gli occhi all'imparare la meccanica quantistica, *aggiustamento post-natale del progetto*. Dopo la

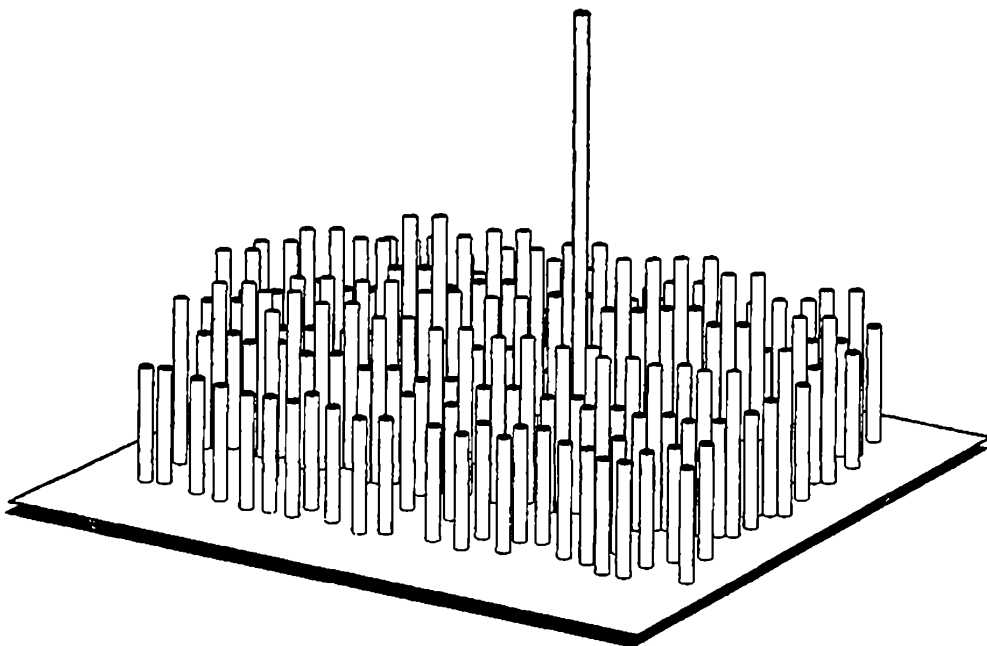
nascita, c'è ancora spazio per dei cambiamenti, che infine, grazie ad un qualche processo, vengano stabilizzati per il resto della vita in un elemento relativamente permanente del progetto (una volta che abbiamo imparato ad andare in bicicletta o a parlare russo, tendiamo a conservare questa capacità).

Come può realizzarsi questo processo di aggiustamento post-natale del progetto? Solo in un modo (non miracoloso): tramite un processo fortemente analogo a quello che fissa il progetto pre-natale, o, in altre parole, un processo di evoluzione per selezione naturale che si verifica dentro l'individuo (dentro il fenotipo). Qualcosa di già fissato nell'individuo tramite l'ordinaria selezione naturale deve giocare il ruolo del selettore meccanico e altre cose devono giocare il ruolo della moltitudine di candidati per la selezione. Molte teorie differenti del processo sono state proposte, ma tutte – tranne quelle semplicemente strampalate o francamente misteriose – presentano la stessa struttura e si differenziano solo per i dettagli dei meccanismi proposti. Per molti anni nel corso del ventesimo secolo, la teoria più autorevole è stata il comportamentismo di B.F. Skinner, nel quale le coppie stimolo-risposta costituivano i candidati per la selezione e gli stimoli di «rinforzo» il meccanismo di selezione. Il ruolo degli stimoli dolorosi e piacevoli – il bastone e la carota – nel plasmare il comportamento è innegabile, ma il meccanismo fondamentale del comportamentismo, il «condizionamento operante», è stato da molti ritenuto troppo semplice per spiegare le complessità dell'aggiustamento post-natale del progetto in specie così sofisticate come quella umana (e probabilmente anche nei piccioni, ma questo è un altro problema). Oggi, l'attenzione si concentra su varie teorie che spostano il processo evoluzionistico dentro il cervello (Dennett, 1974). Svariate versioni di questa idea sono circolate per decenni, e ora, con la possibilità di saggiare i modelli rivali in enormi simulazioni su sistemi di calcolo, ci troviamo di fronte ad una notevole conflittualità, dalla quale faremo bene a stare alla larga.[6]

Per quanto basta ai nostri scopi diciamo solamente che, in un modo o nell'altro, il cervello plastico è capace di riorganizzare se stesso in modo adattativo in risposta alle particolari novità incontrate nell'ambiente dell'organismo, e che il processo tramite il quale il cervello riesce a fare ciò è quasi certamente un processo meccanico fortemente analogo alla selezione naturale. Questo è il primo nuovo mezzo di evoluzione: l'aggiustamento post-natale del progetto nei cervelli individuali. I candidati alla selezione sono le varie strutture cere-

brali che controllano o influenzano i comportamenti e la selezione è realizzata tramite un qualche processo eliminatorio meccanico che è esso stesso geneticamente installato nel sistema nervoso.

Sorprendentemente, questa capacità, essa stessa un prodotto dell'evoluzione genetica tramite selezione naturale, non solo dà all'organismo che la possiede un vantaggio sui suoi cugini cablati rigidamente che non possono riprogettare se stessi, ma si riflette sul processo dell'evoluzione genetica e lo *accelera*. Questo è un fenomeno conosciuto sotto vari nomi ma è meglio conosciuto come Effetto Baldwin (vedi Richards, 1987; Schull, 1990). Ecco come funziona.

Si consideri una popolazione di una specie particolare nella quale è presente una considerevole variazione alla nascita del modo in cui i cervelli sono cablati. Solo uno dei possibili cablaggi, supponiamo, dota il suo possessore di un Buon Trucco – un talento comportamentale che lo protegge o accresce enormemente le sue possibilità. Possiamo rappresentare questo fatto tramite ciò che viene chiamato un panorama adattativo; l'altitudine rappresenta l'adeguatezza (più alto è migliore) e la latitudine e longitudine rappresentano le variabili nel cablaggio (non abbiamo bisogno di specificarle per questo esperimento mentale).

7.1

Come risulta chiaro dalla figura, solo un cablaggio è favorito; gli altri, indipendentemente dalla loro «vicinanza» al buon cablaggio, sono su per giù di uguale adeguatezza tra loro. Questo ago nel pagliaio può essere praticamente invisibile alla

selezione naturale. Anche se pochi fortunati individui sono cablati in questa maniera, la probabilità che la loro fortuna si diffonda nella popolazione delle successive generazioni può essere estremamente ridotta, *a meno che* non ci sia plasticità nel progetto degli individui.

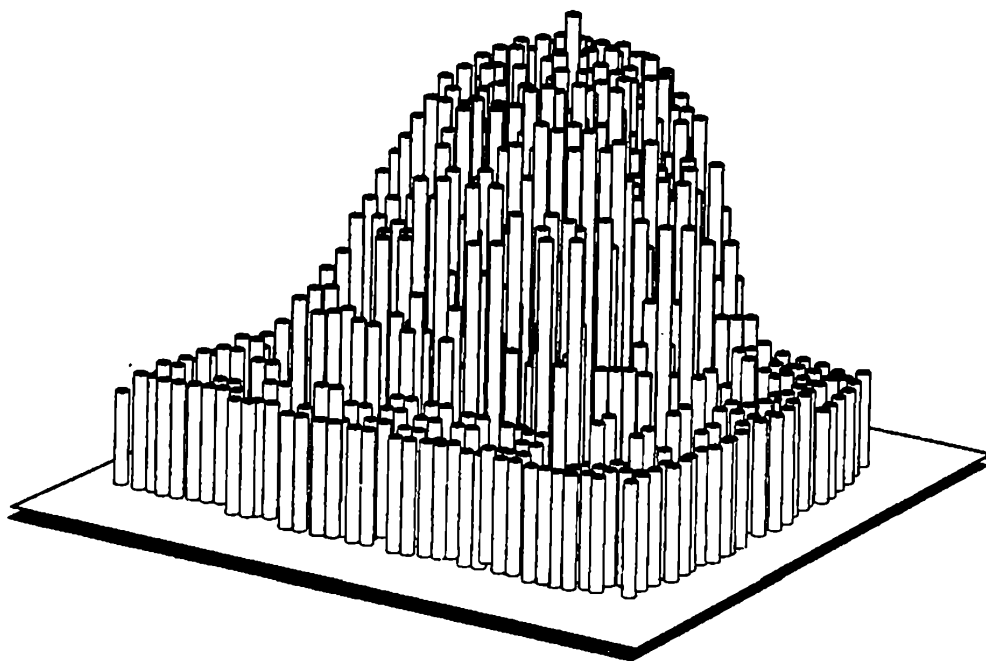
Supponiamo, allora, che tutti gli individui al nastro di partenza siano, geneticamente, differenti, ma che nel corso della loro vita vaghino nello spazio delle possibilità progettuali disponibili, grazie alla loro plasticità. A causa delle particolari circostanze dell'ambiente, tendono tutti a gravitare attorno all'unico cablaggio favorito. C'è un Buon Trucco da imparare nell'ambiente e tutti tendono ad impararlo. Supponiamo che si tratti di un buon trucco che mette in serio svantaggio quelli che non lo imparano mai, e supponiamo che quelli che non lo imparano mai siano quelli che nascono con progetti che si trovano molto lontano nello spazio progettuale (e quindi richiedono una maggiore riprogettazione post-natale) da quelli vicini al Buon Trucco.

Un esempio fantasioso (tratto da Hinton e Nowlan, 1987) ci aiuterà a visualizzare ciò. Supponiamo che ci siano dieci punti nel cervello di ogni animale in cui un «filo» possa essere collegato in uno dei due modi, A o B. Supponiamo che il Buon Trucco sia possibile con il progetto che presenta la cablatura AAABBBAAAA, e che tutte le altre cablature siano pressappoco ugualmente inefficaci per il comportamento. Poiché tutte queste connessioni sono plastiche, ogni animale, nel corso della sua vita, può provare ciascuna delle 2^{10} differenti combinazioni di A e B. Gli animali che sono nati con BAABBBAAAA sono solo un bit lontani dal Buon Trucco (benché possano, naturalmente, allontanarsene provando prima una serie di altre ricablature). Gli altri, che partono con un cablaggio BBBAAABBBB devono compiere almeno dieci ricablature (senza mai ritornare indietro con una cablatura errata) prima di trovare il Buon Trucco. Gli animali che partono con un cervello più vicino all'obiettivo avranno un vantaggio nella lotta per la sopravvivenza su quelli che partono con un cervello più lontano, anche se non c'è nessun *altro* vantaggio selettivo nell'essere nati con una struttura «quasi giusta» rispetto ad una «del tutto errata» (come la figura 7.1 chiarifica). La popolazione della generazione successiva tenderà allora a consistere principalmente di individui più vicini all'obiettivo (e quindi più atti a trovare l'obiettivo nelle loro vite), e questo processo può andare avanti fino a che l'intera popolazione non sia *geneticamente*

stabilizzata sul Buon Trucco. Un Buon Trucco «scoperto» dagli individui in questa maniera può essere passato, in modo relativamente veloce, alle generazioni future.

Se concediamo agli individui una qualche possibilità di imbattersi nel Buon Trucco (e «riconoscerlo» e «aggrapparvisi») nel corso della loro vita, l'ago nel pagliaio che era quasi invisibile (figura 7.1) diventa la sommità di una collina piuttosto evidente che la selezione naturale può scalare (figura 7.2). Questo processo, l'Effetto Baldwin, potrebbe sembrare dapprincipio simile alla discreditata idea lamarckiana della trasmissione genetica dei caratteri acquisiti, ma non è così. Nulla di quello che l'individuo ha appreso viene trasmesso alla sua progenie. Si afferma soltanto che quegli individui abbastanza fortunati da trovarsi più vicini nello spazio dell'esplorazione del progetto al Buon Trucco apprendibile tenderanno ad avere una progenie più numerosa, la quale tenderà a sua volta ad essere più vicina al Buon Trucco. Col succedersi delle generazioni la competizione si fa più serrata, e alla fine chi non è nato con il Buon Trucco (o abbastanza vicino ad esso) non è più in grado di competere. Se non fosse per la plasticità, comunque, l'effetto non si verificherebbe, perché «un capello vale un chilometro», a *meno che* non si possano provare tutte le variazioni fino a trovare quella giusta.

Grazie all'Effetto Baldwin si può dire allora che le specie saggiano in anticipo l'efficacia di particolari progetti differenti grazie all'esplorazione fenotipica (individuale) dello spazio delle possibilità prossime. Se una particolare mossa vincente

7.2

viene così scoperta, essa *creerà* una nuova pressione selettiva: gli organismi che sono più vicini nel panorama adattativo alla scoperta avranno un chiaro vantaggio su quelli più distanti. Ciò significa che le specie dotate di plasticità *tenderanno* ad evolvere più velocemente (e con più «lungimiranza»). Quindi l'evoluzione nel secondo mezzo, la plasticità fenotipica, può accelerare l'evoluzione nel primo mezzo, la variazione genetica. (Vedremo presto l'effetto controbilanciante che sorge come risultato delle interazioni con il terzo mezzo.)

4. *La plasticità nel cervello umano: l'allestimento del palcoscenico*

> Così anche l'intelletto con la sua forza innata si fa degli strumenti intellettuali con i quali si acquista altre forze per altre opere intellettuali e da queste opere si forma altri strumenti, ossia il potere d'indagare ulteriormente; e così avanza gradatamente fino ad attingere il culmine della sapienza.
>
> BARUCH SPINOZA (1661)

I sistemi nervosi rigidamente cablati sono leggeri, consumano poca energia e vanno bene per gli organismi che devono fronteggiare un ambiente stereotipato utilizzando delle risorse limitate. I cervelli più sofisticati, grazie alla loro plasticità, sono capaci non solo di anticipazioni stereotipate, ma anche di adattarsi a linee di tendenza. Persino l'umile rospo possiede un qualche grado di libertà nelle sue risposte alle novità, modificando lentamente gli schemi delle sue attività per seguire – con un considerevole ritardo temporale – quei cambiamenti nelle caratteristiche del suo ambiente che contano di più per il suo benessere (Ewert, 1987). Nel cervello del rospo, uno schema progettuale per trattare con il mondo evolve ad un passo che è più veloce della selezione naturale di molti ordini di grandezza – con le «generazioni» che durano secondi o minuti, non anni. Ma per un controllo veramente potente, hai bisogno di una macchina anticipatrice che si riprogetta radicalmente in pochi millisecondi, e per questo ti serve un *virtuoso* nella produzione del futuro, un sistema che può pensare in anticipo, evitare le abitudini inveterate, risolvere i problemi prima che si presentino, riconoscere dei messaggeri completamente nuovi del bene e del male. Nonostante tutta la nostra stupidità, noi esseri umani siamo molto meglio equipaggiati per questo compito di

ogni altra specie di auto-controllori, e ciò è possibile grazie ai nostri enormi cervelli. Ma come?

Passiamo in rassegna i nostri progressi. Abbiamo delineato sommariamente una storia dell'evoluzione di un cervello di primate – un singolo filo nella trama multidimensionale della storia degli esseri viventi. Basato su millenni di precedenti sistemi nervosi, tale cervello consiste di un aggregato di circuiti specializzati progettati per eseguire compiti particolari nell'economia degli antenati dei primati: rilevatori di «oggetti che si avvicinano rapidamente» collegati con meccanismi per schivarli, rilevatori di «qualcuno mi sta osservando» collegati con discriminatori di «amico, nemico o cibo» collegati con ulteriori sottoprocedure appropriate. Possiamo aggiungere anche qualche circuito specifico dei primati come quelli per la coordinazione dell'occhio e della mano progettati per raccogliere le bacche o i semi, e altri progettati per afferrare i rami o anche per manipolare oggetti vicino alla faccia (Rizzolati, Gentilucci e Matelli, 1985). Grazie agli occhi mobili e ad un'inclinazione verso l'esplorazione e l'aggiornamento, questi cervelli di primati erano regolarmente inondati di informazioni multimediali (o, come i neuroscienziati direbbero, informazioni multimodali), e quindi a loro si poneva un nuovo problema: il problema di un controllo di livello superiore.

Un problema posto è anche un'opportunità, una porta aperta su una nuova porzione di spazio progettuale. Fino ad ora, possiamo supporre, i sistemi nervosi hanno risolto il problema «Ora, che faccio?» tramite un dosaggio relativamente semplice di azioni scelte in un repertorio strettamente limitato – se non proprio le famose quattro C (combatti, corri, cibati e copula) almeno un leggero ampliamento di esse. Ma ora, con l'accresciuta plasticità funzionale e l'accresciuta disponibilità di informazioni «centralizzate» provenienti da tutti i vari specialisti, il problema di cosa fare dopo ha generato un meta-problema: *a cosa pensare* dopo. Va benissimo equipaggiarsi con una sottoprocedura «Tutti gli uomini in coperta!», ma poi, quando tutti gli uomini sono in coperta, si deve avere un modo per trattare con la folla di volontari. Non dobbiamo aspettarci di trovare un adeguato capitano già pronto (cosa stava facendo fino ad allora?), così i conflitti tra volontari devono risolversi senza l'intervento di un dirigente superiore. (Come abbiamo già visto nell'esempio del sistema immunitario, un'azione organizzata e concertata non deve dipendere sempre dal controllo di un dirigente centrale.) Un modello pionieristico di questo tipo di

processo è quello rappresentato dall'architettura del primo Pandemonio di Oliver Selfridge (1959) nell'Intelligenza Artificiale, in cui molti «folletti» rivaleggiano in parallelo per l'egemonia, e poiché il nome usato da Selfridge per questo tipo di architettura è così appropriato, io lo userò in un senso generico in questo libro per riferirmi ad esso e a tutti i suoi discendenti, diretti e indiretti, come la «pianificazione conflittuale» (Norman e Shallice, 1980; Shallice, 1988) e le reti «Tutto al vincitore» di Ballard e Feldman (1982) e i loro discendenti.

Però la pianificazione conflittuale stile pandemonio, guidata in modo piuttosto diretto dalle caratteristiche presenti dell'ambiente, dà al sistema nervoso una capacità limitata di previsione. Come Odmar Neumann ha ipotizzato che le reazioni di orientamento, originariamente prodotte dalle novità nell'ambiente, possano essere avviate endogenamente (dall'interno), così noi possiamo ipotizzare che sussistesse una pressione a sviluppare un modo più endogeno per risolvere il meta-problema di cosa pensare successivamente, una pressione a creare qualcosa all'interno dotato degli immaginari poteri organizzativi di un capitano.

Si consideri a cosa poteva assomigliare il comportamento del nostro ipotetico antenato primate visto dall'esterno (stiamo posponendo di molto tutti i problemi riguardo a quello che un tale primate doveva provare): un animale capace di apprendere nuovi trucchi e quasi continuamente vigile e sensibile alle novità, ma con una «attenzione a breve termine» che tende ad essere «catturata» dalle caratteristiche distraenti dell'ambiente. Nessun progetto a lungo termine per questo animale, almeno nessun nuovo progetto. (Dobbiamo permettere l'esistenza di sottoprocedure stereotipate di lunga durata cablate geneticamente, come le procedure di costruzione del nido negli uccelli, di costruzione di dighe nei castori, di cattura del cibo in uccelli e scoiattoli.)

Sul sostrato di questo sistema nervoso proviamo ora ad immaginare di costruire una mente più umana, con qualcosa che assomigli a un «flusso di coscienza» capace di mantenere quelle sofisticate «successioni di pensiero» su cui la civiltà umana sembra basarsi. Gli scimpanzé sono i nostri consanguinei più vicini – geneticamente più vicini, in effetti, di quanto gli scimpanzé non siano ai gorilla o agli orango – e si pensa comunemente che con loro abbiamo diviso un comune antenato circa sei milioni di anni fa. Dopo questa importante separazione, i nostri cervelli hanno iniziato a svilupparsi in modo

drammaticamente diverso, ma soprattutto nelle dimensioni piuttosto che nella struttura. Mentre gli scimpanzé hanno cervelli approssimativamente delle stesse dimensioni del nostro comune antenato (ed è importante – e difficile – tenere presente che anche gli scimpanzé si sono in qualche modo evoluti rispetto ad esso), i cervelli dei nostri antenati ominidi sono diventati quattro volte più grandi. Questo incremento di volume non si è verificato immediatamente; per vari milioni di anni dopo la separazione dai proto-scimpanzé, i nostri antenati ominidi sono andati avanti con cervelli di dimensioni simili a quelli delle scimmie, malgrado siano diventati bipedi almeno tre milioni e mezzo di anni fa. Poi, quando le età glaciali iniziarono, circa due milioni e mezzo di anni fa, la Grande Encefalizzazione cominciò e la si può considerare essenzialmente completata 150.000 anni fa – *prima* dello sviluppo del linguaggio, dell'agricoltura e dell'abitudine a cuocere i cibi. Perché i cervelli dei nostri antenati siano cresciuti così tanto in così poco tempo (nella scala temporale dell'evoluzione è più un'esplosione che una fioritura) è un problema alquanto dibattuto (per un approccio illuminante, vedi i libri di William Calvin). Non ci sono controversie, invece, sulla natura del prodotto di questa evoluzione: il cervello dei primi *Homo sapiens* (che vissero da circa 150.000 anni fa alla fine della più recente glaciazione solo 10.000 anni fa) era un cervello enormemente complesso dotato di ineguagliabile plasticità e difficilmente distinguibile dal nostro in forma e dimensione. Questo è importante: la sorprendente crescita del cervello degli ominidi fu essenzialmente completata *prima* dello sviluppo del linguaggio, e quindi non può essere la risposta alle complessità della mente rese possibili dal linguaggio. Le specializzazioni innate per il linguaggio, ipotizzate dal linguista Noam Chomsky ed altri, e che ora iniziano ad essere confermate nei dettagli della neuroanatomia, costituiscono un'aggiunta *molto* recente e affrettata, sicuramente un'utilizzazione di preesistenti circuiti di sequenzializzazione (Calvin, 1989a) accelerata dall'Effetto Baldwin. Inoltre, l'espansione più considerevole dei poteri *mentali* umani (così come è testimoniata dallo sviluppo dell'agricoltura, dell'arte, della cucina o, in una parola, della civiltà) si è verificata anche più recentemente, dalla fine dell'ultima era glaciale, in un baleno di 10.000 anni che corrisponde ad un istante dalla prospettiva evoluzionistica che misura le tendenze in milioni di anni. I nostri cervelli sono equipaggiati alla nascita con quasi gli stessi poteri di quelli dei cervelli dei nostri antenati di 10.000

anni fa. Così gli enormi progressi dell'*Homo sapiens* negli ultimi 10.000 anni devono essere quasi tutti dovuti alla capacità di imbrigliare la plasticità di quel cervello in un modo radicalmente nuovo – creando qualcosa che assomiglia ad un *software* per potenziare i soggiacenti poteri (Dennett, 1986).

In breve, i nostri antenati devono aver appreso alcuni Buoni Trucchi che si potevano fare con i loro *hardware* riadattabili, e la nostra specie ha solo cominciato a trasferirli, tramite l'Effetto Baldwin, nel genoma. Inoltre, come vedremo, ci sono ragioni per credere che malgrado la pressione selettiva iniziale a favore di una graduale «cablatura rigida» di questi Buoni Trucchi, essi hanno così alterato la natura dell'ambiente per la nostra specie che non c'è più stata nessuna significante pressione selettiva che richiedesse un'ulteriore cablatura rigida. È come se quasi tutta la pressione selettiva sullo sviluppo del progetto del sistema nervoso umano sia stata travolta dagli effetti collaterali di questa nuova opportunità del progetto sfruttata dai nostri antenati.

Fino a questo punto io ho evitato di parlare di *rappresentazioni* di aspetti del mondo esterno nei più semplici sistemi nervosi. Si può pensare che i vari progetti che abbiamo considerato, sia quelli plastici sia quelli cablati rigidamente, siano sensibili a (o rispondenti a, o progettati con un occhio verso, o utilizzino informazioni su) varie caratteristiche dell'ambiente dell'organismo, e quindi in un senso molto limitato si può parlare di rappresentazioni. Ora, comunque, è bene fermarci un attimo a considerare quali caratteristiche di tali complessi progetti debbano condurci a considerarli come sistemi di rappresentazioni.

Alcuni degli aspetti variabili del cervello sono necessari semplicemente per fornire un mezzo alle configurazioni continuamente mutevoli delle attività cerebrali che in qualche modo *registrano* o in ogni caso *seguono* le caratteristiche variabili importanti dell'ambiente. Qualcosa nel cervello deve cambiare per poter seguire l'uccello che vola o l'abbassamento della temperatura dell'aria o un mutamento interno nello stato dell'organismo stesso – la caduta dello zucchero nel sangue, l'aumento di anidride carbonica nei polmoni. Inoltre – e questo è ciò che ci porta nel campo delle autentiche rappresentazioni – queste mutevoli configurazioni interne devono essere in grado di continuare a «seguire» (in senso ampio) le caratteristiche a cui si riferiscono anche quando sono temporaneamente tagliate fuori dal rapporto causale con i loro referenti. «Una zebra che ha

avvistato un leone non dimentica dove sta il leone se smette di guardarlo per un momento. Il leone non dimentica dove sta la zebra» (Margolis, 1987, p. 53). Si faccia il paragone con il fenomeno più semplice del girasole che segue il passaggio del sole attraverso il cielo, aggiustando il suo angolo come un pannello solare mobile per massimizzare la luce solare che cade su di esso. Se il sole viene momentaneamente oscurato, il girasole non può tracciare la sua traiettoria; il meccanismo che è sensibile al passaggio del sole non rappresenta il passaggio del sole in questo senso ampio. Le reali rappresentazioni si cominciano a trovare in molti animali inferiori (e non dovremmo escludere, *a priori*, la possibilità di reale rappresentazione nelle piante), ma negli esseri umani la capacità di rappresentazione è schizzata alle stelle.

Tra le cose che un cervello umano adulto può in qualche modo rappresentare ci sono non solo:

1. la posizione del corpo e degli arti
2. una macchia di luce rossa
3. un livello di fame
4. un livello di sete
5. l'aroma di un eccellente vino rosso

ma anche:

6. l'aroma di un eccellente vino rosso *come* l'aroma di uno Chambertin 1971
7. Parigi
8. Atlantide
9. la radice quadrata del più grande numero primo inferiore a 20
10. il concetto di una combinazione tra un apriscatole e un cavatappi placcata al nichel.

Sembra quasi certo che il cervello di nessun altro animale possa rappresentare 6-10 e inoltre che sia necessario un considerevole processo di aggiustamento del cervello umano infantile prima che esse possano essere registrate o rappresentate. Le prime cinque, al contrario, potrebbero ben essere delle cose che quasi ogni cervello può (in qualche senso) rappresentare senza allenamento.

In ogni caso, il modo in cui il cervello rappresenta la fame deve in qualche modo differire, fisicamente, dal modo in cui

rappresenta la sete – poiché deve dirigere un comportamento differente che dipende da ciò che viene rappresentato. Ci deve anche essere, all'altro estremo, una differenza tra il modo in cui un particolare cervello umano adulto rappresenta Parigi e quello in cui rappresenta Atlantide, giacché pensare ad una non è pensare all'altra. Come può un particolare stato o evento nel cervello rappresentare una caratteristica del mondo piuttosto di un'altra?[7] E qualunque cosa sia ciò che fa in modo che una caratteristica del cervello rappresenti ciò che rappresenta, *come fa* a rappresentare ciò che rappresenta? Qui ancora una volta (ho paura che questo ritornello finirà per diventare noioso!) c'è un largo spettro di possibilità, stabilite dai processi evoluzionistici: alcuni elementi del sistema di rappresentazione possono essere – in realtà devono essere (Dennett, 1969) – innati, il resto deve essere «appreso». Mentre alcune delle categorie che sono fondamentali nella vita (come la sete e la fame) sono senza dubbio «date» a noi tramite il modo in cui siamo cablati alla nascita, altre devono essere sviluppate per conto nostro.[8]

Come riusciamo a farlo? Probabilmente tramite un processo di generazione e selezione di configurazioni di attività neurale nella corteccia cerebrale, l'ampio mantello convoluto che è rapidamente germogliato nel cranio umano e che ora copre completamente il sottostante cervello animale più antico. Dire semplicemente che è un processo evolutivo che avviene principalmente nella corteccia lascia troppe cose avvolte nel mistero, e a questo livello di complessità e sofisticazione, anche se riuscissimo a spiegare il processo al livello delle sinapsi o dei fasci di neuroni, rimarremmo nella confusione rispetto ad altri aspetti di ciò che deve avvenire. Se vogliamo trarre un qualche senso da tutto ciò, dobbiamo prima sollevarci ad un livello più generale e astratto, e, una volta comprese le linee generali del processo di livello superiore, possiamo pensare di scendere di nuovo ad un livello più meccanico del cervello.

La plasticità rende possibile l'apprendimento, ma è molto meglio se da qualche parte lì nell'ambiente c'è *qualcosa da apprendere* che è già il prodotto di un precedente processo progettuale, così che ognuno di noi non deve reinventare la ruota. L'evoluzione culturale, e la trasmissione dei suoi prodotti, è l'altro nuovo mezzo di evoluzione e dipende dalla plasticità fenotipica in un modo molto simile a quello in cui la plasticità fenotipica dipende dalla variazione genetica. Noi esseri umani abbiamo usato la nostra plasticità non solo per apprendere, ma per apprendere come apprendere meglio, e poi

abbiamo appreso meglio come apprendere meglio come apprendere meglio, e così via. Abbiamo anche appreso come fare in modo che i frutti di questo apprendimento siano disponibili ai novizi. Noi in qualche modo *installiamo* un *sistema* di abitudini già inventato e ampiamente revisionato («debugged») nel cervello parzialmente non strutturato.

5. *L'invenzione di buone e cattive abitudini di autostimolazione*

> Come posso dire che cosa penso finché non vedo che cosa dico?
>
> E.M. FORSTER (1960)

> Noi parliamo, non solo per dire agli altri cosa pensiamo, ma anche per dire a noi stessi cosa pensiamo.
>
> J. HUGHLINGS JACKSON (1915)

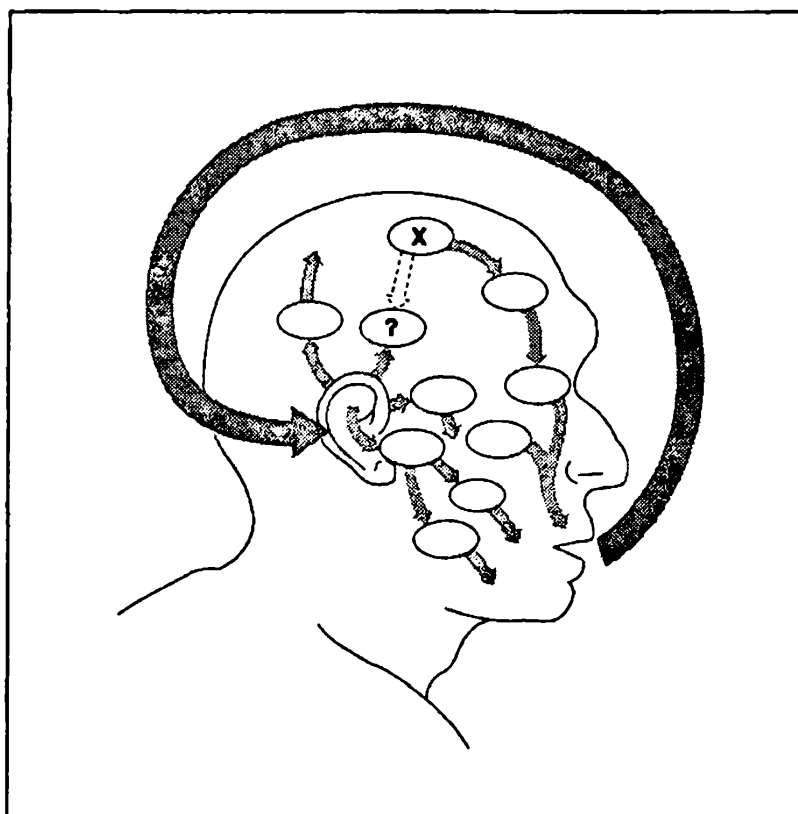
Come si è arrivati a questa condivisione di software? Una storia alla Kipling ci indicherà una strada possibile. Si consideri un tempo nella storia dei primi *Homo sapiens* quando il linguaggio – o forse dovremmo chiamarlo protolinguaggio – cominciava appena a svilupparsi. Questi nostri antenati erano dei bipedi onnivori che vivevano in gruppi piuttosto piccoli di consanguinei e che probabilmente avevano sviluppato delle abitudini di vocalizzazione finalizzate a scopi specifici, simili a quelle degli scimpanzé e dei gorilla o anche di specie molto più lontanamente correlate come i cercopitechi (Cheney e Seyfarth, 1990). Possiamo supporre che gli atti comunicativi (o quasi-comunicativi) eseguiti tramite queste vocalizzazioni non fossero ancora degli atti linguistici completamente sviluppati (Bennett, 1976), in cui l'intenzione del Parlante di raggiungere un certo effetto nell'Uditorio dipende dal riconoscimento da parte dell'Uditorio di quella stessa intenzione.[9] Ma possiamo supporre che questi nostri antenati, come i primati contemporanei in grado di vocalizzare, discriminassero tra differenti parlanti e uditori in occasioni differenti, utilizzando le informazioni su quello che entrambe le parti potrebbero credere o volere.[10] Per esempio, l'ominide Alf non si sforzerebbe di far credere all'ominide Bob che non c'è cibo nella caverna (grugnendo «No-cibo-qui») se Alf credesse che Bob già sapesse che non c'è cibo nella caverna. E se Bob pensasse che Alf lo voglia ingannare, Bob considererebbe la vocalizzazione di Alf con cauto scetticismo.[11]

Ora accadeva talvolta, possiamo congetturare, che quando uno di questi ominidi si trovava in difficoltà con un progetto «chiedesse aiuto», e in particolare «chiedesse informazioni». Talvolta l'uditorio presente poteva rispondere «comunicando» qualcosa che aveva l'effetto giusto sul richiedente, facendolo uscire dalle sue abitudini erronee o facendogli «vedere» una soluzione al problema. Affinché questa pratica potesse prendere piede nella comunità, era necessario che il richiedente fosse in grado di contraccambiare, al momento opportuno, nel ruolo di rispondente. Essi dovevano essere dotati della capacità comportamentale di sentirsi spinti a pronunciare occasionalmente qualcosa di «utile» quando sottoposti alla «richiesta» di altri. Per esempio, se un ominide conosceva qualcosa e gli veniva rivolta una «domanda» su di essa, ciò avrebbe avuto l'effetto normale, ma sicuramente non privo di eccezioni, di spingerlo a «dire ciò che sapeva».

In altre parole, sto proponendo che c'era un tempo nell'evoluzione del linguaggio in cui le vocalizzazioni avevano la funzione di ricavare e condividere utili informazioni, ma non si deve presupporre che uno spirito cooperativo di aiuto reciproco avrebbe un valore per la sopravvivenza o sarebbe un sistema stabile se emergesse. (Vedi, per esempio, Dawkins, 1982 pp. 55 sg.; vedi anche Sperber e Wilson, 1986.) Invece, dobbiamo presupporre che i costi e i benefici dovuti alla partecipazione a tale pratica fossero in qualche modo «visibili» a quelle creature, e che un numero sufficiente di esse vide che i benefici per *se stessi* superavano i costi e così tali abitudini comunicative si affermarono nella comunità.

Poi un bel giorno (in questa ricostruzione teorica), uno di questi ominidi chiese aiuto «erroneamente» quando non c'era nessun uditorio utilizzabile a portata di voce – tranne lui stesso! Quando udì la sua propria richiesta, la stimolazione lo spinse a produrre proprio quel tipo di vocalizzazione in aiuto di altri che la richiesta da parte degli altri avrebbe provocato. E con grande gioia, quella creatura scoprì che aveva spinto se stesso a rispondere alla sua propria domanda.

Tramite questo esperimento mentale volutamente molto semplificato sto solo tentando di giustificare l'affermazione che la pratica di porre domande a se stessi possa sorgere come naturale effetto collaterale del porre domande agli altri, e che la sua utilità sarebbe similare: sarebbe un comportamento di cui si potrebbe scoprire la capacità di migliorare le proprie possibilità grazie ad un controllo dell'azione meglio informato. Affinché

una pratica del genere sia utile è necessario che le preesistenti relazioni di accesso *nel* cervello dell'individuo non siano ottimali. Si deve supporre, in altre parole, che benché *all'interno del cervello* ci sia già la giusta informazione per qualche scopo, essa sia in mano allo specialista sbagliato; il sottosistema nel cervello che ha bisogno dell'informazione non può ottenerla direttamente dallo specialista – perché l'evoluzione semplicemente non ha avuto il tempo di fornire un tale «filo». Spingere lo specialista a «mandare in onda» l'informazione nell'ambiente, comunque, e poi affidarsi ad un paio di orecchie esistenti (e un sistema uditivo) per captarla, costituirebbe un modo per costruire un «filo virtuale» tra i sottosistemi pertinenti.[12]

7.3

Un tale atto di autostimolazione potrebbe aprire la strada a nuove preziose connessioni tra le componenti interne del cervello. Per dirla in modo rozzo, spingere un'informazione attraverso le orecchie e il sistema uditivo di qualcuno può servire a stimolare proprio il giusto tipo di collegamento che egli stava cercando, può far scattare proprio il giusto meccanismo associativo, pungolare proprio il giusto bocconcino mentale fin sulla punta della sua lingua. Si può, allora, dire qualcosa, udirsi dirla e poi ottenere la risposta desiderata.

Una volta che abitudini ancora rozze di autostimolazione vocale iniziano ad affermarsi come Buoni Trucchi nel compor-

tamento di una popolazione di ominidi, dovremmo aspettarci che esse vengano rapidamente raffinate, sia tramite le abitudini comportamentali apprese della popolazione, sia, grazie all'Effetto Baldwin, tramite le predisposizioni genetiche e gli ulteriori potenziamenti dell'efficienza e dell'efficacia. In particolare, possiamo congetturare che vengano riconosciute le speciali virtù del parlare a se stesso *sotto voce*, che conducono successivamente ad un parlare a se stesso del tutto silenzioso. Il processo silenzioso manterrebbe il circuito autostimolativo, ma si disfarebbe delle parti auditive e di vocalizzazione periferica, che non erano essenziali. Questa innovazione avrebbe l'effetto ulteriore, opportunisticamente rafforzato, di raggiungere una certa privatezza della pratica di autostimolazione cognitiva. (Nel prossimo capitolo, considereremo come queste linee abbreviate di comunicazione potrebbero funzionare.) Una tale privatezza sarebbe particolarmente utile nei casi in cui a portata di voce ci sono dei cospecifici in grado di comprendere. Questo comportamento privato di parlare a se stesso potrebbe benissimo non essere il miglior modo immaginabile per emendare l'architettura funzionale esistente del proprio cervello, ma sarebbe comunque un miglioramento facilmente rinvenibile e a portata di mano – e potrebbe essere più che sufficiente. Sarebbe lento e laborioso in paragone ai processi cognitivi veloci e inconsci su cui è basato, giacché deve far uso di larghi tratti del sistema nervoso «progettati per altri scopi» – in particolare per la produzione e la comprensione del discorso udibile. Sarebbe altrettanto lineare (limitato ad un argomento alla volta) della comunicazione sociale da cui è evoluto. E dipenderebbe, almeno al principio, dalle categorie informazionali incorporate nelle azioni che sfrutta. (Se ci fossero solo cinquanta cose che un ominide può «dire» a un altro, ci sarebbero solo cinquanta cose che egli potrebbe dire a se stesso.)

Parlare ad alta voce è solo una possibilità. Fare disegni per se stessi è un altro atto di automanipolazione facilmente apprezzabile. Supponiamo che un giorno uno di questi ominidi abbia tracciato indolentemente due linee parallele sul pavimento della sua caverna, e che quando guardò che cosa aveva fatto, queste due linee gli abbiano ricordato, visivamente, i bordi paralleli del fiume che avrebbe dovuto attraversare poche ore più tardi, e che ciò gli abbia ricordato di prendere con sé la corda di liane per attraversarlo. Se non avesse tracciato quel «disegno», possiamo supporre, avrebbe camminato fino al fiume e *poi* capito, dopo una rapida ispezione visiva, che aveva bisogno della sua

corda, e sarebbe dovuto ritornare indietro. Questo notevole risparmio di tempo e di energia potrebbe costituire un buon motivo per promuovere una nuova abitudine, e per raffinarla successivamente in un *privato* «far disegni» per «l'occhio della propria mente».

Il talento umano per inventare nuovi percorsi di comunicazione interna si mostra talvolta in modo particolarmente vivido nei casi di danni cerebrali. La gente si dimostra straordinariamente brava nel superare i danni cerebrali e non si tratta *mai* di «guarire» o riparare i circuiti danneggiati. Vengono scoperti invece nuovi modi per eseguire i vecchi trucchi, e questo grazie soprattutto al ruolo che l'esplorazione attiva gioca nella riabilitazione. Un aneddoto particolarmente suggestivo viene dalla ricerca su pazienti con cervello diviso (Gazzaniga, 1978). Gli emisferi destro e sinistro sono normalmente connessi da un ampio ponte di fibre chiamato *corpus callosum*. Quando esso viene chirurgicamente reciso (nel trattamento di gravi epilessie), i due emisferi perdono i loro principali «fili» diretti di interconnessione e sono praticamente non comunicanti. Se ad un tale paziente si chiede di identificare un oggetto – come una matita – infilando la mano in una borsa e toccandolo, il successo dipende da quale mano sta usando. La maggior parte della cablatura del corpo è strutturata *controlateralmente*, con l'emisfero sinistro che ottiene le sue informazioni da – e controlla – la parte destra del corpo, e viceversa. Poiché l'emisfero sinistro controlla normalmente il linguaggio, quando il paziente infila nella borsa la mano destra, può dire facilmente cosa c'è dentro, ma se infila la mano sinistra, solo l'emisfero destro ottiene l'informazione che l'oggetto è una matita, e non riesce ad esprimere linguisticamente questo fatto. Ma talvolta, sembra, l'emisfero destro scopre un astuto stratagemma: trova la punta della matita, la spinge contro la palma, e fa in modo che un segnale di acuto dolore risalga dalla mano sinistra verso il cervello. Poiché alcune fibre del dolore non sono controlateralizzate, l'emisfero sinistro, che controlla il linguaggio, ha un suggerimento: è qualcosa di sufficientemente aguzzo da provocare dolore: «È aguzzo – forse è una penna, una matita?». L'emisfero destro, cogliendo questa vocalizzazione può offrire alcuni suggerimenti – aggrottando le ciglia su *penna* e ridendo su *matita* – così tramite un rapido giro di «Lascia o raddoppia?» l'emisfero sinistro viene condotto alla risposta giusta. Esistono non pochi aneddoti su tali ingegnosi espedienti inventati là per là dai pazienti con cervello diviso, ma dovremmo trattarli con

prudenza. Potrebbero essere quello che sembrano: casi che evidenziano la destrezza con cui il cervello può scoprire e realizzare strategie autostimolatorie per migliorare le comunicazioni interne in assenza della «desiderata» cablatura. Ma potrebbero essere anche fantasie inconsapevolmente abbellite di ricercatori che sperano proprio di ottenere simili «prove». Questo è il problema con gli aneddoti.

Potremmo divertirci a immaginare altri possibili scenari per «l'invenzione» di utili modi di autostimolazione, ma rischieremmo di oscurare il fatto che non tutte le invenzioni del genere sarebbero utili per la sopravvivenza. Una volta che una generica abitudine di autostimolazione esploratoria si sia inculcata in qualche modo o modi, può benissimo generare una schiera di variazioni non funzionali (ma non particolarmente controproducenti). Ci sono, dopo tutto, molte varietà di autostimolazione che presumibilmente non hanno alcun effetto positivo sulla cognizione o sul controllo, ma che, per le solite ragioni darwiniane, non subiscono l'estinzione e possono essere sospinte verso una cristallizzazione (culturale *o* genetica) in determinate sottopopolazioni. Candidati verosimili sono il dipingersi di blu, l'autoflagellarsi con una verga, l'incidersi delle figure sulla pelle, il lasciarsi morire di fame, il dirsi ripetutamente una formula «magica» e l'osservarsi l'ombelico. Se queste pratiche rappresentano delle abitudini che vale la pena di assumere, i loro pregi come potenziatori di adattamento non sono così «ovvi» da averle spinte in qualche predisposizione genetica conosciuta, ma forse si tratta di invenzioni troppo recenti.

I vari tipi di autostimolazione che potenziano l'organizzazione cognitiva sono ora probabilmente in parte innati e in parte appresi e specifici dell'individuo. Proprio come qualcuno può notare che strofinandosi in una certa maniera può produrre determinati effetti collaterali desiderabili che sono controllabili solo parzialmente e indirettamente – e quindi può dedicare tempo e attenzione per sviluppare ed esplorare le tecniche per produrre quegli effetti collaterali – così può semi-consciamente esplorare le tecniche di autostimolazione cognitiva, sviluppando uno stile personale con particolari pregi o difetti. Alcune persone sono più brave di altre ed alcune non imparano mai i trucchi, ma molto è dovuto alla condivisione e all'insegnamento. La *trasmissione* culturale, lasciando arrivare quasi tutti al Buon Trucco, può appiattire la sommità della collina dell'adeguatezza (figura 7.2), creando un altipiano o un pianoro che

diminuisce la pressione selettiva per spostare il Trucco nel genoma. Se quasi tutti diventano sufficientemente bravi nell'arrangiarsi nel mondo civilizzato, la pressione selettiva per spostare il Buon Trucco nel genoma si estingue o perlomeno perde intensità.

6. *Il terzo processo evoluzionistico: memi ed evoluzione culturale*[13]

Come abbiamo appreso a mungere le vacche e poi ad addomesticarle per il nostro tornaconto, così abbiamo imparato a mungere le menti, nostre e degli altri, in determinate maniere; ed ora le tecniche di auto e reciproca stimolazione sono profondamente impiantate nella nostra cultura e nella nostra educazione. Il modo in cui la cultura è diventata un mezzo per conservare e trasmettere le innovazioni (non soltanto quelle che riguardano la coscienza) è importante per comprendere le origini della struttura della coscienza umana, giacché costituisce un altro mezzo di evoluzione.

Uno dei passi sicuramente più importanti che il cervello umano intraprende nel grandioso processo di auto-progettazione post-natale è sicuramente quello tramite il quale si adatta alle condizioni locali che contano di più: esso si trasforma rapidamente (nel giro di due o tre anni) in un cervello swahili, giapponese o inglese. Che passo! È come salire su un razzo!

Non è importante per i nostri fini se questo processo è chiamato apprendimento o sviluppo differenziale; esso avviene così rapidamente e così naturalmente da non lasciar dubbi che il genotipo umano includa molti adattamenti che servono specificamente a facilitare l'acquisizione del linguaggio. Tutto ciò è avvenuto molto rapidamente, in termini evoluzionistici, ma questo è proprio quanto ci aspetteremmo, dato l'Effetto Baldwin. Essere in grado di parlare è proprio uno di quei Buoni Trucchi che mettono in serio svantaggio chi è lento a raggiungere i blocchi di partenza. I nostri primi antenati che hanno iniziato a parlare hanno quasi sicuramente avuto un bel da fare per riuscirci, ma noi siamo i discendenti dei virtuosi che c'erano tra loro.[14]

Una volta che i nostri cervelli hanno costruito le strade di entrata e di uscita per i veicoli del linguaggio, essi vengono rapidamente *parassitati* (in senso letterale, come vedremo) da entità che si sono evolute per prosperare proprio in tali nicchie: i *memi*. Le linee fondamentali della teoria dell'evoluzione per

selezione naturale sono chiare: l'evoluzione si verifica ogni volta che siano presenti le seguenti condizioni:

1. variazione: una continua abbondanza di elementi differenti
2. eredità o replicazione: gli elementi hanno la capacità di creare copie o repliche di se stessi
3. «adeguatezza» differenziata: il numero di copie di un elemento che vengono create in un dato momento varia, a seconda delle interazioni tra le caratteristiche di quell'elemento (qualunque cosa sia ciò che lo rende differente dagli altri elementi) e le caratteristiche dell'ambiente in cui continuare ad esistere.

Si noti che in questa definizione, benché sia tratta dalla biologia, non si dice specificatamente nulla su molecole organiche, nutrimento o perfino vita. È una caratterizzazione molto generale ed astratta dell'evoluzione per selezione naturale. Come lo zoologo Richard Dawkins ha sottolineato, il principio fondamentale è

> che ogni genere di vita si evolve mediante la sopravvivenza differenziata di entità che si replicano [...]
>
> Il gene, la molecola di DNA, è l'entità replicante che si trova a predominare sul nostra pianeta. Ce ne possono essere altre. Se ci sono, purché siano soddisfatte certe altre condizioni, esse tenderanno quasi inevitabilmente a diventare la base di un processo evolutivo.
>
> Ma è proprio necessario andare su mondi lontani per trovare altri generi di replicanti e, di conseguenza, altri generi di evoluzione? Io credo che proprio su questo pianeta sia venuto recentemente alla luce un genere nuovo di replicante. L'abbiamo sotto gli occhi: è ancora in una fase infantile, si muove goffamente qua e là nel suo brodo primordiale, ma sta già conoscendo cambiamenti evolutivi a una velocità tale che il vecchio gene ansimante gli resta parecchio indietro. (1976, p. 206)

Questi nuovi replicanti sono, grosso modo, le idee. Non le «idee semplici» di Locke e Hume (l'idea del rosso o l'idea del cerchio o del caldo o del freddo), ma quel tipo di idee complesse che si strutturano in distinte unità degne di essere memorizzate – come le idee di

ruota
indossare vestiti
vendetta

triangolo retto
alfabeto
calendario
l'*Odissea*
calcolo
scacchi
disegno prospettico
evoluzione per selezione naturale
impressionismo
la tarantella
il decostruzionismo

Intuitivamente queste sono delle unità culturali più o meno identificabili, ma possiamo dire qualcosa di più preciso sul modo in cui tracciamo i confini – sul perché «Re-Fa♯-La» non è un'unità e lo è invece il tema dell'adagio della Settima Sinfonia di Beethoven: le unità sono gli elementi più piccoli che replicano se stessi con affidabilità e fecondità. Dawkins conia un termine per tali unità: *memi* –

unità di trasmissione culturale o unità di *imitazione*. «Mimema» potrebbe andare, perché deriva da una radice greca appropriata, ma io voglio un bisillabo che suoni un po' come «gene» [...] lo si può vedere come imparentato con «memoria» o con la parola francese *même*.

Esempi di memi sono le melodie, le idee, gli slogan e i modi di dire, le mode dell'abbigliamento, le tecniche per fabbricare vasi o per costruire archi. Come i geni si propagano nel fondo comune dei geni passando da un corpo all'altro con gli spermatozoi o gli ovuli, così i memi si propagano nel fondo comune dei memi passando da un cervello all'altro con un processo che, in senso lato, si può chiamare imitazione. Se uno scienziato legge o sente parlare di una buona idea, la trasmette ai suoi colleghi e ai suoi studenti, la menziona nei suoi articoli e nelle sue lezioni. Se l'idea attecchisce, si può dire che essa si propaga da sola diffondendosi da un cervello all'altro. (1976)

Ne *Il gene egoista*, Dawkins ci invita a prendere letteralmente l'idea di evoluzione dei memi: essa non è semplicemente analoga all'evoluzione biologica o genetica, non è soltanto un processo che può essere descritto metaforicamente nel linguaggio evoluzionistico, ma è un fenomeno che obbedisce esattamente alle leggi della selezione naturale. La teoria dell'evoluzione per selezione naturale si mantiene neutrale rispetto alle differenze tra memi e geni; si tratta semplicemente di differenti

tipi di replicatori che si evolvono in mezzi differenti e a ritmi differenti. E come i geni degli animali non poterono apparire su questo pianeta fino a che l'evoluzione delle piante non aprì loro la strada (creando un'atmosfera ricca di ossigeno e una grande quantità di sostanze nutritive convertibili), così l'evoluzione dei memi non poté iniziare fino a che l'evoluzione degli animali non aprì loro la strada creando una specie – *Homo sapiens* – con cervelli in cui si potevano rifugiare e abitudini di comunicazione che potevano utilizzare per propagarsi.

Questo è un nuovo modo di pensare alle idee. È anche, spero di far vedere, un buon modo; ma offre una prospettiva che dapprincipio sembra sicuramente sconvolgente, se non terrificante. Possiamo riassumerla nella frase:

Uno studioso è soltanto un modo in cui una biblioteca crea un'altra biblioteca.

Non so tu, ma io non sono inizialmente attratto dall'idea che il mio cervello sia solo un mucchio di letame in cui le larve delle idee delle altre persone si rinnovano, prima che delle loro copie siano spedite in una diaspora informazionale. Sembra che la mia mente sia privata della sua importanza di autore e critico. Chi comanda, secondo questa prospettiva – noi o i nostri memi?

Non c'è, naturalmente, una risposta semplice e questo fatto sta al cuore della confusione che circonda l'idea di un *sé*. La coscienza umana è in gran parte un prodotto non solo della selezione naturale ma anche dell'evoluzione culturale. Il modo migliore per capire il contributo che i memi danno alla creazione delle nostre menti è quello di seguire da vicino i passi tipici del pensiero evoluzionistico.

La prima regola valida per i memi, così come per i geni, è che la replica non avviene necessariamente per il bene di qualcuno; i replicatori fioriscono per il bene [...] della replicazione – quale che ne sia la ragione. Come Dawkins si esprime,

Un meme che obbliga l'organismo in cui si trova a precipitare giù da una scogliera avrà lo stesso destino di un gene che si comporta in modo analogo: esso tenderà a venir eliminato dal fondo comune dei memi. [...] Questo non significa che il criterio ultimo per il successo della selezione dei memi sia la sopravvivenza del gene. [...] Ovviamente un meme che provoca il suicidio dell'individuo che lo ospita ha gravi svantaggi, ma non necessariamente fatali. [...] Un meme per il suicidio può benissimo diffondersi, come ad esempio nel caso in cui un

drammatico e ben pubblicizzato martirio spinga altri a morire per una causa profondamente sentita, e questi nuovi martiri ne ispirano altri e così via. (1982, pp. 141-143)

Il punto importante è che non c'è una connessione *necessaria* tra il potere replicativo del meme, la sua «adeguatezza» dal *suo* punto di vista, e il suo contributo alla *nostra* adeguatezza (quale che sia lo standard con cui la giudichiamo). La situazione non è però totalmente disperata; benché alcuni memi ci spingono sicuramente a collaborare alla loro replica *malgrado* li giudichiamo inutili o brutti o perfino pericolosi per la nostra salute e il nostro benessere, molti – la maggior parte, se siamo fortunati – dei memi che si replicano lo fanno non solo *con* la nostra benedizione, ma *grazie* alla nostra stima. Credo che si possa affermare tranquillamente che alcuni memi sono, tutto sommato, buoni *dalla nostra prospettiva*, e non solo dalla loro prospettiva di auto-replicatori egoisti: mi riferisco a memi molto generali come la cooperazione, la musica, la scrittura, l'educazione, la consapevolezza ambientale, la riduzione degli armamenti; e memi particolari come *Le nozze di Figaro, Moby Dick*, i contenitori riciclabili, i trattati SALT. Altri memi sono più controversi; possiamo capire perché si diffondono e perché dopo tutto dobbiamo tollerarli malgrado i problemi che ci creano: i centri commerciali, i fast food, la pubblicità in televisione. Altri ancora sono sicuramente dannosi, ma estremamente difficili da sradicare: l'antisemitismo, il terrorismo, i virus dei calcolatori, le scritte che imbrattano i muri.

I geni sono invisibili; sono trasportati dai veicoli dei geni (gli organismi) nei quali tendono a produrre gli effetti caratteristici (effetti «fenotipici») che, alla lunga, determinano i loro destini. Anche i memi sono invisibili, e sono trasportati dai veicoli dei memi – immagini, libri, frasi (in particolari linguaggi, scritte o orali, sulla carta o supporti magnetici, ecc.). Anche gli strumenti, gli edifici e le invenzioni in generale sono veicoli dei memi: un carro con le ruote non trasporta solo il grano o il suo carico da un luogo all'altro, trasporta anche la brillante idea di un carro con le ruote da una mente all'altra. L'esistenza di un meme dipende dalla sua realizzazione fisica in qualche mezzo; se tutte le sue realizzazioni fisiche vengono distrutte, quel meme si è estinto. Potrebbe, naturalmente, fare una successiva riapparizione indipendente – proprio come i geni dei dinosauri potrebbero, in linea di principio, riaggregarsi in un lontano futuro, ma il dinosauro che creerebbero e abiterebbero non sarebbe il

discendente degli originali dinosauri, o perlomeno non più direttamente di quanto lo siamo noi. Il destino dei memi – l'eventualità che le loro copie o le copie delle loro copie seguitino a esistere e a riprodursi – dipende dalle forze selettive che agiscono direttamente sui veicoli fisici che li realizzano.

I veicoli dei memi popolano il nostro mondo assieme a tutta la flora e la fauna, grande e piccola. Ma generalmente parlando sono «visibili» solo dalla specie umana. Si consideri l'ambiente di un piccione medio nella città di New York, ogni giorno un numero più o meno uguale di parole, immagini e altri segni e simboli assale sia i suoi occhi e le sue orecchie che gli occhi e le orecchie di ogni abitatore umano di quella città. Questi veicoli fisici dei memi possono influire profondamente sul benessere del piccione, ma non in virtù del meme che trasportano – per il piccione non cambia nulla se la briciola che trova sta sotto una pagina del «National Enquirer» o del «New York Times».

Per gli esseri umani, d'altra parte, ogni veicolo di meme è un potenziale amico o nemico, trasporta un dono che aumenterà i nostri poteri o un inutile balocco che ci distrarrà, appesantirà la nostra memoria e scombussolerà i nostri giudizi. Possiamo paragonare questi invasori aerotrasportati che penetrano nei nostri occhi e nelle nostre orecchie con i parassiti che entrano nel nostro corpo per altre vie: ci sono i parassiti benefici come i batteri nel nostro stomaco, senza i quali non potremmo digerire il cibo; i parassiti tollerabili, che non vale la pena di eliminare (tutti gli abitatori della nostra pelle e dei nostri capelli, per esempio); e gli invasori pericolosi che sono difficili da sradicare (il virus dell'AIDS, per esempio).

Fino qui, potrà sembrare che la prospettiva dal punto di vista del meme sia soltanto un modo pittoresco per riorganizzare varie osservazioni ben conosciute sul modo in cui gli aspetti della nostra cultura ci influenzano, e si influenzano tra loro. Ma Dawkins osserva che nelle nostre spiegazioni noi tendiamo a trascurare il fatto che «un tratto culturale può essersi evoluto in una data maniera semplicemente perché è *vantaggioso per se stesso*» (1976, p. 214). Questa è la chiave per rispondere alla domanda se il meme mème sia uno di quelli che dovremmo sfruttare e replicare. Secondo la prospettiva normale, le seguenti proposizioni sono virtualmente tautologiche:

L'idea X è stata creduta dalla gente perché X è stata ritenuta vera.
La gente ha approvato X perché la gente ha trovato X bella.

Ciò che richiede una speciale attenzione è il caso in cui, malgrado la verità o la bellezza di un'idea, essa *non* è accettata, o malgrado la sua bruttezza o falsità *è* accettata. La prospettiva dal punto di vista del meme si dimostra utile quando vogliamo spiegare queste deviazioni; ciò che è tautologico per essa è

Il meme X si è diffuso tra la gente perché X è un buon replicatore.

Ora, tra le due prospettive esiste una correlazione non aleatoria; e non è un caso: noi non sopravviveremmo se non avessimo una abitudine migliore del semplice caso a scegliere i memi che ci aiutano. I nostri sistemi meme-immunitari non sono a prova di errore, ma nemmeno irrimediabilmente difettosi. Possiamo basarci, come rozza regola generale, sulla coincidenza tra le due prospettive: generalmente, i memi buoni sono quelli che sono anche buoni replicatori.

La teoria diventa interessante solo quando prendiamo in esame le eccezioni, i casi in cui c'e una divergenza tra le due prospettive; solo se la teoria dei memi ci permette di comprendere meglio le deviazioni dallo schema normale, essa avrà qualche speranza di essere accettata. (Nota che, nei suoi propri termini, la replica più o meno riuscita del meme meme è strettamente indipendente dai suoi pregi epistemologici; si può diffondere malgrado la sua pericolosità o estinguersi malgrado le sue virtù.)

I memi ora si diffondono nel mondo alla velocità della luce e si replicano ad un ritmo che fa sembrare statici perfino i moscerini e le cellule di lievito. Saltano promiscuamente da veicolo a veicolo, da mezzo a mezzo, ed è praticamente impossibile tenerli in quarantena. I memi, come i geni, sono *potenzialmente* immortali, ma, come i geni, dipendono dall'esistenza di una catena continua di veicoli fisici, che si conservano malgrado la seconda legge della termodinamica. I libri sono relativamente stabili e le iscrizioni sui monumenti lo sono ancora di più, ma tutto ciò, a meno che non sia sotto la protezione di un conservatore umano, tende a dissolversi col tempo. Come per i geni, l'immortalità è più un fatto di replicazione che di longevità dei veicoli individuali. La conservazione dei memi platonici, attraverso una serie di copie di copie, è un esempio particolarmente sorprendente di ciò. Benché recentemente siano stati trovati dei testi platonici su frammenti di papiro più o meno contemporanei a Platone, la sopravvivenza dei memi

platonici non deve quasi nulla ad una tale persistenza a così lungo raggio. Le biblioteche odierne contengono migliaia se non milioni di copie fisiche (e traduzioni) del *Menone* di Platone e i capostipiti della trasmissione di questo testo sono da secoli ridotti in polvere.

La semplice replica fisica dei veicoli non è sufficiente ad assicurare la longevità dei memi. Alcune migliaia di copie rilegate di un nuovo libro possono scomparire senza lasciare traccia nel giro di pochi anni e chissà quante brillanti lettere al direttore, riprodotte in centinaia di migliaia di copie, scompaiono nelle discariche e negli inceneritori ogni giorno? Potrà arrivare il giorno in cui un valutatore non umano di memi sarà sufficiente a selezionare e ordinare i memi per la conservazione, ma per il momento, i memi devono ancora far affidamento almeno indirettamente al fatto che uno o più dei loro veicoli passino almeno un breve stadio pupale in un tipo speciale di nicchia per memi: la mente umana.

La quantità di menti è limitata e ogni mente ha una capacità limitata di contenere memi, quindi c'è una considerevole competizione tra i memi per entrare in quante più menti possibili. Tale competizione è la maggior forza selettiva nella memosfera e, proprio come nella biosfera, la sfida è stata ingaggiata con grande ingegnosità. Per esempio, quali che siano i pregi (dalla nostra prospettiva) dei seguenti memi, essi hanno in comune la proprietà di avere espressioni fenotipiche che tendono a rendere più probabile la riuscita delle loro repliche disarmando o invalidando le forze ambientali che tenderebbero ad estinguerli: il meme per *fede*, che scoraggia l'esercizio di quella sorta di giudizio critico che potrebbe far decidere che l'idea di fede è tutto sommato un'idea dannosa (Dawkins, 1976, p. 212); il meme per *tolleranza* o *libertà di parola*; il meme di includere in una catena di sant'Antonio l'avvertimento sul terribile destino di quelli che nel passato hanno rotto la catena; il meme della *teoria della cospirazione*, che ha una risposta già pronta all'obiezione che non ci sono prove evidenti della cospirazione: «Certo che no, questo ti dimostra quanto la cospirazione sia potente!». Forse alcuni di questi memi sono «buoni» e altri «cattivi»; ciò che hanno in comune è un effetto fenotipico che tende sistematicamente a disarmare le forze selettive schierate contro di loro. A parità di condizioni, la memetica delle popolazioni prevede che i memi della teoria della cospirazione persisteranno indipendentemente dalla loro verità e che il meme per fede è adatto ad assicurare la propria sopravvivenza,

e quella dei memi religiosi ad esso collegati, anche negli ambienti più razionalistici. In realtà, il meme per fede mostra una *adeguatezza dipendente dalla frequenza*: prospera meglio quando è messo in minoranza da memi razionalistici; in un ambiente con pochi scettici, il meme per fede tende a svanire per il disuso.

Anche altri concetti si possono trasferire facilmente dalla genetica delle popolazioni; ecco un caso di ciò che un genetista chiamerebbe *posizione vincolata*: due memi a cui capita di essere fisicamente legati insieme così che tendono sempre a replicarsi insieme, cosa che influenza le loro possibilità. Esiste una magnifica marcia cerimoniale, familiare a molti di noi, una marcia che sarebbe molto usata per le inaugurazioni, i matrimoni e altre occasioni festive, e che forse porterebbe quasi all'estinzione sia *Pomp and Circumstance* che la Marcia nuziale del *Lohengrin*, se non fosse per il fatto che il suo meme musicale è troppo strettamente legato al meme del suo titolo, che molti di noi tendono a ricordare non appena ascoltano la musica: l'inutilizzabile capolavoro di Sir Arthur Sullivan *Ecco Sua Eccellenza il Boia!*

Il paradiso che tutti i memi cercano di raggiungere è la mente umana, ma la mente umana è essa stessa un artefatto creato quando i memi ristrutturano un cervello umano per renderlo un habitat a loro più confacente. Le vie di ingresso e di uscita sono modificate per adattarsi alle situazioni locali e sono rafforzate da vari congegni artificiali che potenziano la fedeltà e la prolissità di replicazione: una mente di madrelingua cinese differisce drammaticamente da una di madrelingua francese e una mente alfabetizzata differisce da una analfabeta. I memi contraccambiano gli organismi in cui risiedono con un'incalcolabile quantità di vantaggi – e qualche cavallo di Troia gettato lì per precauzione, senza dubbio. I cervelli umani normali non sono tutti uguali; differiscono sensibilmente per dimensioni, forma e per gli innumerevoli dettagli di connessione da cui dipendono i loro talenti. Ma le differenze più sorprendenti nei talenti dei cervelli umani dipendono dalle differenze microstrutturali indotte dai vari memi che vi sono penetrati e vi hanno preso residenza. I memi si rafforzano vicendevolmente: il meme per educazione, per esempio, è un meme che rinforza lo stesso processo di implantazione dei memi.

Ma se è vero che le menti umane sono esse stesse in larga parte una creazione dei memi, allora non possiamo rimanere fermi alla polarità della situazione prospettata all'inizio; non si tratta di «memi contro noi», perché le precedenti infestazioni di

memi hanno già giocato un ruolo fondamentale nel determinare *chi* o *che cosa siamo*. La mente «indipendente» che lotta per proteggersi dai memi esterni e pericolosi è un mito; c'è, nel sottofondo, una persistente tensione tra l'imperativo biologico dei geni e gli imperativi dei memi, ma sarebbe sciocco «schierarsi» con i nostri geni – significherebbe commettere l'errore più grossolano della sociobiologia del senso comune. Su quali fondamenta possiamo, allora, basarci mentre lottiamo per mantenerci in equilibrio nella tempesta di memi in cui siamo irretiti? Se non è la replicazione, qual è l'ideale eterno rispetto al quale «noi» giudicheremo il valore dei memi? Dovremmo notare che i memi per i concetti normativi – per *dovere* e *bene* e *verità* e *bellezza* – sono tra i più resistenti cittadini della nostra mente e che tra i memi che ci costituiscono giocano un ruolo centrale. La nostra esistenza come noi stessi, ciò che siamo come pensatori – e non come organismi – non è indipendente da questi memi.

Riassumendo: l'evoluzione dei memi ha il potere di contribuire considerevolmente al potenziamento del progetto del sottostante meccanismo del cervello – a grande velocità, se paragonato al passo lento con cui Madre Natura affronta i compiti di ricerca e sviluppo genetici. La discreditata idea lamarckiana della trasmissione genetica dei caratteri acquisiti individualmente poteva inizialmente piacere ai biologi anche per la sua presunta capacità di catapultare le nuove invenzioni nel genoma, accelerando così l'evoluzione. (Per un'elegante demolizione del lamarckismo, vedi la discussione di Dawkins ne *Il fenotipo esteso*, 1982.) Questo non succede, non può succedere. L'Effetto Baldwin accelera davvero l'evoluzione, favorendo l'inserimento nel genoma di Buoni Trucchi scoperti individualmente, tramite la strada indiretta della creazione di nuove pressioni selettive a seguito della diffusa accettazione dei Buoni Trucchi da parte degli individui. Ma l'evoluzione culturale, che avviene ancora più velocemente, permette agli individui di acquisire, tramite la trasmissione culturale, dei Buoni Trucchi che sono stati affinati da predecessori che non erano neppure loro antenati genetici. Gli effetti di tale condivisione di buoni progetti sono così potenti che l'evoluzione culturale ha probabilmente cancellato quasi tutte le delicate pressioni dell'Effetto Baldwin. I miglioramenti del progetto che si ricevono dalla propria cultura – raramente si deve «reinventare la ruota» – schiacciano probabilmente la maggior parte delle differenze genetiche nel progetto del cervello, eliminando i vantaggi di quelli che partono leggermente avvantaggiati alla nascita.

Tutti e tre i mezzi – l'evoluzione genetica, la plasticità fenotipica e l'evoluzione memetica – hanno contribuito al progetto della coscienza umana, ognuno a suo tempo e a una velocità crescente. Se confrontata con la plasticità fenotipica, che esiste da milioni di anni, l'evoluzione memetica vera e propria rappresenta un fenomeno estremamente recente, essendo diventata una forza potente solo negli ultimi centomila anni ed esplodendo con lo sviluppo della civiltà meno di diecimila anni fa. Essa è ristretta ad una specie, *Homo sapiens*, e possiamo notare che ci ha portato alle soglie di un ulteriore quarto mezzo di ricerca e sviluppo potenziale, grazie ai memi della scienza: la diretta revisione del sistema nervoso individuale tramite l'ingegneria neuroscientifica e la revisione del genoma tramite l'ingegneria genetica.

7. *I memi della coscienza: la macchina virtuale da installare*

> Benché sia possibile che un organo non sia stato originariamente formato per qualche scopo speciale, se ora serve a tal fine noi siamo autorizzati a dire che è stato inventato proprio per quello. Basandoci sullo stesso principio, se un uomo dovesse costruire una macchina per qualche scopo speciale, usando però vecchi ingranaggi, molle e pulegge, solo leggermente modificati, si potrebbe dire che l'intera macchina, con tutte le sue parti, è stata inventata proprio per quello scopo. Così da un capo all'altro della natura quasi ogni parte di ogni essere vivente è servita probabilmente, in una condizione leggermente modificata, per scopi diversi e ha funzionato nel meccanismo vivente in forme desuete e differenti.
>
> CHARLES DARWIN, 1874

> Un vasto cervello, come un vasto governo, potrebbe non essere in grado di fare cose semplici in modo semplice.
>
> DONALD HEBB, 1958

> La spinta più poderosa nello sviluppo dell'uomo è il piacere che prova nella sua abilità. Gli piace far bene ciò che fa e, avendolo fatto bene, gli piace farlo meglio.
>
> JACOB BRONOWSKI, 1973

Nella mia ipotetica ricostruzione storica ho sottolineato il fatto

che i nostri antenati, come noi, provavano piacere nell'autoesplorarsi in vari modi relativamente non guidati – stimolandosi sempre di più e vedendo poi cosa succedeva. Tenendo conto della plasticità del cervello, associata a un'inquietudine e una curiosità innate che ci conducono a esplorare ogni angolo del nostro ambiente (di cui i nostri corpi sono un elemento così importante e onnipresente), non ci sorprende, retrospettivamente, che ci imbattessimo in strategie di autostimolazione o automanipolazione che condussero all'assunzione di abitudini e disposizioni che modificarono radicalmente la struttura comunicativa interna dei nostri cervelli e che divennero parte della cultura – memi – e quindi disponibili a tutti.

La trasformazione di un cervello umano causata dalle infestazioni di memi costituisce una modificazione profonda delle competenze di quell'organo. Come abbiamo già notato, le differenze tra un cervello di madrelingua cinese e uno di madrelingua inglese sarebbero sufficienti a spiegare le enormi differenze riscontrabili nelle competenze dei due cervelli, differenze immediatamente riconoscibili nel comportamento e altamente significanti in molti contesti sperimentali. Ci si ricordi, per esempio, di quanto è importante negli esperimenti con soggetti umani che lo sperimentatore (l'eterofenomenologo) sappia se il soggetto ha capito o no le istruzioni. Queste differenze funzionali, benché presumibilmente siano tutte realizzate fisicamente tramite delle configurazioni di cambiamenti microscopici nel cervello, sono del tutto invisibili agli occhi dei neuroscienziati, ora e probabilmente per sempre, cosicché se vogliamo avere un'idea dell'architettura funzionale *creata* da tali infestazioni di memi, dovremmo trovare un livello più alto di descrizione. Fortunatamente, un tale livello può essere ricavato dall'informatica. Il livello di descrizione e spiegazione di cui abbiamo bisogno è *analogo* (ma non identico) a uno dei «livelli software» di descrizione dei calcolatori: ciò che dobbiamo capire è come la coscienza umana possa essere realizzata grazie al funzionamento di una *macchina virtuale* creata dai memi nel cervello.

Ecco l'ipotesi che difenderò:

La coscienza umana è essa stessa un enorme complesso di memi (o più esattamente, di effetti provocati dai memi nel cervello) che si può comprendere egregiamente pensando al funzionamento di una macchina virtuale «*neumanniana*» implementata sull'architettura parallela di un cervello che non era progettato

per attività del genere. I poteri di questa *macchina virtuale* accrescono notevolmente i sottostanti poteri dell'*hardware* su cui gira, ma nello stesso tempo molte delle sue caratteristiche più strane, e soprattutto delle sue limitazioni, possono essere spiegate come prodotti collaterali dei *kludge* che rendono possibile questa strana ma efficace riutilizzazione di un organo già esistente per nuovi scopi.

Il senso di questa ipotesi emergerà presto con chiarezza dal gergo involuto in cui l'ho espressa. Perché ho usato il gergo? Perché questi termini ci permettono di esprimere importanti concetti cui solo recentemente le persone che riflettono sulla mente hanno potuto attingere. Nessun'altra parola esprime questi concetti così nettamente e vale la pena conoscerli adeguatamente. Così, con l'aiuto di una breve digressione storica, li introdurrò e li porrò nel contesto in cui li useremo.

Due dei più importanti inventori dei calcolatori furono il matematico inglese Alan Turing e il fisico e matematico ungaro-americano John von Neumann. Benché Turing abbia avuto una grande esperienza pratica diretta progettando e costruendo le macchine elettroniche appositamente studiate per decifrare i codici segreti che aiutarono gli alleati a vincere la seconda guerra mondiale, è al suo lavoro teorico esclusivamente astratto, e cioè allo sviluppo del suo concetto di *Macchina di Turing Universale*, che dobbiamo l'avvio dell'Era dei Calcolatori. Von Neumann capì che era possibile trasformare l'astrazione di Turing (che era in realtà «filosofica» – un esperimento mentale e non una proposta ingegneristica) in qualcosa di abbastanza concreto da essere considerato il progetto (ancora piuttosto astratto) di un vero e proprio calcolatore elettronico. Quel progetto astratto, conosciuto come l'*architettura di von Neumann*, può essere rintracciato in quasi tutti i calcolatori del mondo al giorno d'oggi, dal gigantesco *mainframe* al *chip* che sta al cuore del più modesto personal computer.

Un calcolatore ha un'architettura base *fissa* o *cablata rigidamente* (l'*hardware*), ma dotata di grande plasticità grazie alla *memoria*, che può immagazzinare sia i *programmi* (altrimenti conosciuti come *software*) sia i *dati*, le configurazioni puramente momentanee che servono a prendere nota di tutto ciò che si deve rappresentare. I calcolatori, come i cervelli, sono quindi solo incompletamente progettati alla nascita e dispongono di una flessibilità che può essere usata come mezzo per creare architetture più specificatamente disciplinate, macchine con

compiti speciali, ognuna con un suo modo per ricevere gli stimoli dell'ambiente – tramite la tastiera o altri dispositivi di ingresso (*input devices*) – e produrre poi le risposte – tramite uno schermo CRT (tubo a raggi catodici) o altri dispositivi di uscita (*output devices*).

Queste strutture temporanee sono «fatte di regole piuttosto che di cavi», e gli informatici le chiamano *macchine virtuali*.[15] Una macchina virtuale è ciò che ottieni quando imponi un particolare insieme di regole (più letteralmente: disposizioni o regolarità di transizione) su tutta questa plasticità. Considera qualcuno a cui abbiano ingessato un braccio fratturato. Il gesso limita notevolmente i movimenti del braccio e condiziona anche quelli del resto del corpo a causa del suo peso e delle sue dimensioni. Ora considera un mimo (diciamo, Marcel Marceau) che imita qualcuno con un braccio ingessato; se il mimo è bravo, i movimenti del suo corpo saranno limitati nello stesso modo; egli ha un *gesso virtuale* al braccio – che sarà «quasi visibile». Chiunque abbia una certa familiarità con un sistema di videoscrittura o *word processor* conosce almeno una macchina virtuale; e se ha usato vari programmi differenti di videoscrittura oppure ha usato un foglio elettronico o utilizzato per vari videogiochi il medesimo calcolatore con cui scrive, conosce varie macchine virtuali, che esistono a rotazione su una particolare macchina reale. Le differenze sono ben chiare e l'utente sa sempre con quale macchina virtuale sta interagendo ogni volta.

Tutti sanno che programmi differenti conferiscono differenti poteri ai calcolatori, ma non tutti conoscono i dettagli. Alcuni di essi sono importanti per la nostra storia, così devo chiedere indulgenza e fornire una breve ed elementare illustrazione del processo inventato da Alan Turing.

Turing non stava cercando di inventare la videoscrittura o i videogiochi quando fece le sue interessanti scoperte. Stava pensando, autocoscientemente e introspettivamente, al modo in cui egli, un matematico, procedeva quando risolveva un problema matematico o eseguiva una *computazione*, e prese l'importante decisione metodologica di provare a suddividere la *sequenza* dei suoi *atti mentali* nei suoi componenti primitivi. «Cosa faccio», deve essersi chiesto, «quando eseguo un calcolo? Be', prima mi chiedo quale regola devo applicare, e poi applico la regola, e poi scrivo il risultato, e poi leggo il risultato, e poi mi chiedo che cosa devo fare successivamente, e poi...». Turing era un pensatore estremamente ben organizzato, ma il suo flusso di coscienza, come il tuo o il mio o quello di James Joyce, era senza

dubbio un variegato guazzabuglio di immagini, decisioni, impressioni, ricordi, ecc., dal quale egli riuscì a distillare l'essenza matematica: lo scheletro, la sequenza minima delle operazioni che potevano raggiungere gli obiettivi che egli raggiungeva nelle lussureggianti e tortuose attività della sua mente cosciente. Ne risultò la caratterizzazione di ciò che noi oggi chiamiamo una macchina di Turing, una brillante idealizzazione e semplificazione di un fenomeno iperrazionale e iperintellettuale: un matematico che esegue un calcolo rigoroso. L'idea fondamentale è composta di cinque elementi:

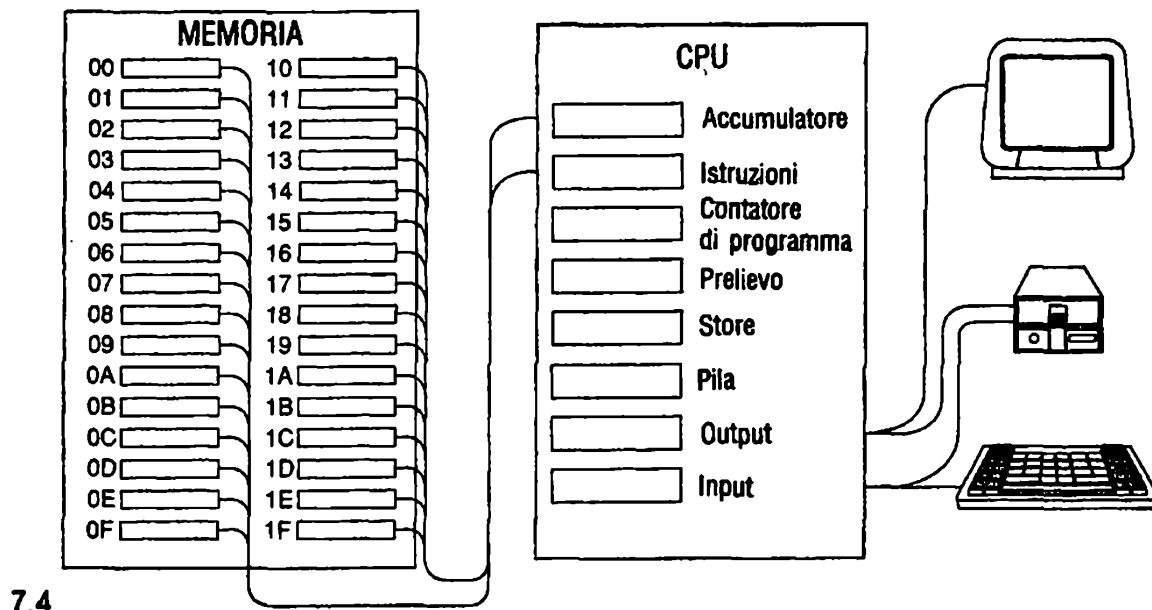
1. un processo *seriale* (gli eventi si verificano uno alla volta), in
2. un *ambiente di lavoro* estremamente limitato, in cui
3. sia i *dati* che le *istruzioni* sono portati da
4. una *memoria* inerte ma altamente affidabile,
5. su cui si svolgono un insieme finito di *operazioni primitive.*

Nella formulazione originale di Turing, l'ambiente di lavoro è costituito da una testina analizzatrice che scorre su un nastro di carta suddiviso in caselle per vedere, una casella alla volta, se c'è scritto uno zero o un uno. A seconda di quello che «vede», o cancella lo zero o l'uno e stampa l'altro simbolo, o lascia la casella invariata. Poi si sposta a sinistra o a destra di una casella e guarda di nuovo, guidata in ogni singolo momento da un insieme finito di istruzioni rigidamente assegnate che formano lo *stato di transizione della macchina.* Il nastro di carta costituisce la memoria.

L'insieme delle operazioni primitive immaginate da Turing (gli «atti atomici dell'introspezione», se vuoi) era volutamente povero, cosicché non potevano esserci dubbi sulla loro realizzabilità meccanica. In altre parole, era importante per le finalità matematiche di Turing che fosse indiscutibile che ogni singolo passo del processo che egli stava studiando fosse così semplice, così stupido, che potesse essere eseguito da qualunque sempliciotto – da qualcuno facilmente sostituibile da una macchina: LEGGI, CANCELLA, STAMPA, SPOSTATI A SINISTRA DI UNO SPAZIO e così via.

Egli capì, naturalmente, che la sua caratterizzazione ideale poteva servire, indirettamente, come schema progettuale per una effettiva macchina computazionale, e anche altri lo capirono, in particolare John von Neumann, che modificò le idee fondamentali di Turing per creare l'architettura astratta per il

primo calcolatore numerico praticamente realizzabile. Noi chiamiamo questa architettura la *macchina di von Neumann.*

7.4

A sinistra c'è la *memoria o memoria RAM* (*Random Access Memory* o memoria ad accesso casuale), dove risiedono sia i dati che le istruzioni, codificati in sequenze di cifre binarie o «bit», come 00011011 o 01001110. Il processo seriale di Turing si svolge nell'ambiente di lavoro costituito da due «registri» denominati *accumulatore* e *registro delle istruzioni.* Dopo che un'istruzione viene copiata elettronicamente sul registro delle istruzioni, esso la *esegue.* Per esempio, se l'istruzione (tradotta nel linguaggio naturale) dice «libera l'accumulatore», il calcolatore mette il numero 0 nell'accumulatore, e se l'istruzione dice «somma il contenuto del registro di memoria 07 al numero nell'accumulatore», il calcolatore va' a prendere il numero che sta all'indirizzo 07 (il contenuto può essere qualunque numero) e lo somma al numero nell'accumulatore. E così via. Quali sono le operazioni primitive? Fondamentalmente, le operazioni aritmetiche: somma, sottrazione, moltiplicazione e divisione; le operazioni per spostare i dati: va' a prendere, immagazzina, emetti, immetti; e (il cuore della «logica» dei calcolatori) le istruzioni *condizionali*, quali «SE il numero nell'accumulatore è più grande di zero, ALLORA vai all'istruzione nel registro 29; altrimenti vai alla prossima istruzione». A seconda del modello di calcolatore, ci sono da un minimo di sedici ad alcune centinaia di operazioni primitive, tutte cablate in circuiti specializzati. Ogni operazione primitiva è codificata tramite un'unica configurazione binaria (per es., SOMMA può essere 1011 e SOTTRAI 1101) e ogniqualvolta queste particolari sequenze arrivano al registro delle istruzioni

funzionano come prefissi telefonici che aprono meccanicamente le linee con l'appropriato circuito specializzato – il circuito addizionatore o il circuito sottrattore, e così via. I due registri, in cui possono comparire solo un'istruzione e un valore alla volta, costituiscono la famosa «strozzatura di von Neumann», il luogo angusto in cui tutte le attività del sistema devono passare in fila indiana. In un calcolatore veloce, si possono svolgere milioni di operazioni al secondo che collegate insieme producono gli effetti apparentemente magici che l'utente osserva.

Tutti i calcolatori numerici sono i *diretti discendenti* di questo schema progettuale, e sebbene molte modificazioni e miglioramenti siano stati fatti, come tutti i vertebrati, condividono la stessa architettura sottostante. Le operazioni fondamentali, apparendo così matematiche, non sembrano dapprincipio avere molto in comune con le fondamentali «operazioni» di un normale flusso di coscienza – pensare a Parigi, godersi l'aroma del pane appena uscito dal forno, fantasticare sulle prossime vacanze – ma ciò non preoccupò Turing o von Neumann. Ciò che era importante per loro è che questa sequenza di azioni può «in linea di principio» essere ampliata fino a includere *tutti* «i pensieri razionali» e forse anche tutti «i pensieri irrazionali». Per un'ironia della storia particolarmente significativa, questa architettura fu descritta in modo fuorviante dalla stampa popolare fin dalla sua prima comparsa. Queste nuove ed affascinanti macchine di von Neumann furono chiamate «giganteschi cervelli elettronici», ma esse erano in realtà *gigantesche menti elettroniche*, imitazioni elettroniche – decise semplificazioni – di quel che William James chiamò flusso di coscienza, la tortuosa sequenza di contenuti mentali coscienti meravigliosamente descritta da James Joyce nei suoi romanzi. L'architettura del cervello, al contrario, è massicciamente parallela, con milioni di canali operazionali simultaneamente attivi. Ciò che dobbiamo capire è come un fenomeno seriale joyceano (o, come ho detto, «neumanniano») possa esistere, con tutte le sue familiari peculiarità, nel tumulto parallelo del cervello.

Ecco una cattiva idea: i nostri antenati ominidi avevano bisogno di pensare in un modo più logico, più sofisticato, così la selezione naturale ha gradualmente progettato e installato una macchina di von Neumann cablata rigidamente nell'emisfero sinistro (quello «logico», «cosciente») della corteccia umana. Spero sia chiaro dalla precedente narrazione evoluzionistica che sebbene sia *logicamente* possibile, ciò non è affatto plausibile biologicamente – ai nostri antenati potevano altrettanto facil-

mente essere spuntate le ali o le pistole sulle mani fin dalla nascita. Non è questo il modo in cui l'evoluzione svolge il suo lavoro.

Noi *sappiamo* che c'è qualche cosa anche solo *vagamente* simile alla macchina di von Neumann nel cervello, perché sappiamo di avere delle menti coscienti «per introspezione» e le menti che in questa maniera scopriamo sono simili alle macchine di von Neumann almeno per questo: sono l'ispirazione delle macchine di von Neumann! Questo fatto storico ha lasciato una traccia fossile particolarmente significativa: i programmatori informatici ti diranno che è terribilmente difficile programmare un calcolatore parallelo come quelli che attualmente si stanno sviluppando, e relativamente facile programmare una macchina seriale, di von Neumann. Quando programmi una convenzionale macchina di von Neumann, hai a disposizione un comodo sostegno; se le cose si fanno difficili, ti chiedi «Che cosa farei io se fossi la macchina, per cercare di risolvere il problema?» e questo ti conduce ad una risposta del genere «Bene, prima farei questo, e poi farei quello, ecc.». Ma se ti chiedessi «Che cosa farei in questa situazione se fossi un elaboratore parallelo a mille canali?» non sapresti cosa rispondere; non hai nessuna personale familiarità con – nessun «accesso diretto a» – i processi che si verificano in mille canali contemporaneamente, anche se questo è quello che accade nel tuo cervello. Il tuo unico accesso a quello che avviene nel tuo cervello si presenta in un «formato» sequenziale che ricorda straordinariamente l'architettura di von Neumann – sebbene presentarlo così sia storicamente rovesciato.

C'è una grande differenza, come abbiamo visto, tra un'ordinaria architettura seriale di un calcolatore e l'architettura parallela del cervello. Questo fatto viene spesso addotto come obiezione all'Intelligenza Artificiale, che tenta di creare un'intelligenza simile a quella umana escogitando programmi che girano (quasi sempre) su macchine di von Neumann. Questa differenza di architettura è anche un'importante differenza teorica? In un senso, no. Turing ha dimostrato – e questo è probabilmente il suo contributo più importante – che la sua Macchina di Turing Universale può calcolare qualsiasi funzione che possa essere calcolata da qualunque calcolatore, con qualunque architettura. In effetti la Macchina di Turing Universale rappresenta il perfetto camaleonte matematico, capace di imitare *qualunque* altra macchina calcolante e di fare, nel corso di tale imitazione, *esattamente* le stesse cose di quella

macchina. L'unica cosa che devi fare è fornire alla Macchina di Turing Universale un'appropriata descrizione dell'altra macchina, e – come Marcel Marceau (la macchina imitante universale), armato di una particolareggiata coreografia, procede senz'altro a una perfetta imitazione basata su quella descrizione – essa diviene, virtualmente, l'altra macchina. Un programma di calcolo può allora essere visto o come una lista di istruzioni primitive da seguire o come una descrizione di una macchina da imitare.

Puoi imitare Marcel Marceau che imita un ubriaco che imita un battitore di baseball? Potrai pensare che la cosa più difficile è non perdersi tra i differenti livelli di imitazione, ma per una macchina di von Neumann questo avviene naturalmente. Una volta che hai una macchina di von Neumann a disposizione, puoi costruirci sopra tante macchine virtuali nidificate come scatole cinesi. Per esempio, puoi prima trasformare la tua macchina di von Neumann in una macchina Unix (il sistema operativo Unix) e poi implementare una macchina Lisp (il linguaggio di programmazione Lisp) sulla macchina Unix – assieme a WordStar, Lotus 123, e una schiera di altre macchine virtuali – e poi implementare un calcolatore che gioca a scacchi sulla tua macchina Lisp. Ogni macchina virtuale è riconoscibile dalla sua *interfaccia con l'utente* – il modo in cui appare sullo schermo CRT e il modo in cui risponde all'input – e questa auto-presentazione viene spesso chiamata l'*illusione dell'utente*, poiché l'utente non sa – e non si preoccupa di sapere – come la specifica macchina virtuale che sta usando è implementata sull'hardware. Per lui non è importante se la macchina virtuale è distante uno, due, tre o dieci livelli dall'hardware.[16] (Per esempio qualcuno che utilizza WordStar può riconoscere, e interagire con, la macchina virtuale WordStar ovunque la trovi, indipendentemente dalle differenze dell'hardware sottostante.)

Così una macchina virtuale è un insieme temporaneo di regole altamente strutturate imposte all'hardware sottostante da un *programma*: una ricetta strutturata di centinaia di migliaia di istruzioni che danno all'hardware un insieme vasto e interdipendente di abitudini o disposizioni a reagire. Se ti fermi a guardare i dettagli microscopici di tutte quelle istruzioni che si susseguono turbinosamente nel registro delle istruzioni, perderai di vista la foresta per gli alberi; se fai un passo indietro, comunque, l'architettura funzionale che emerge da tutto quell'arrangiamento microscopico apparirà chiaro: consiste di *cose* virtuali

come blocchi di testo, cursori, gomme per cancellare, pennelli, file

e di *luoghi* virtuali

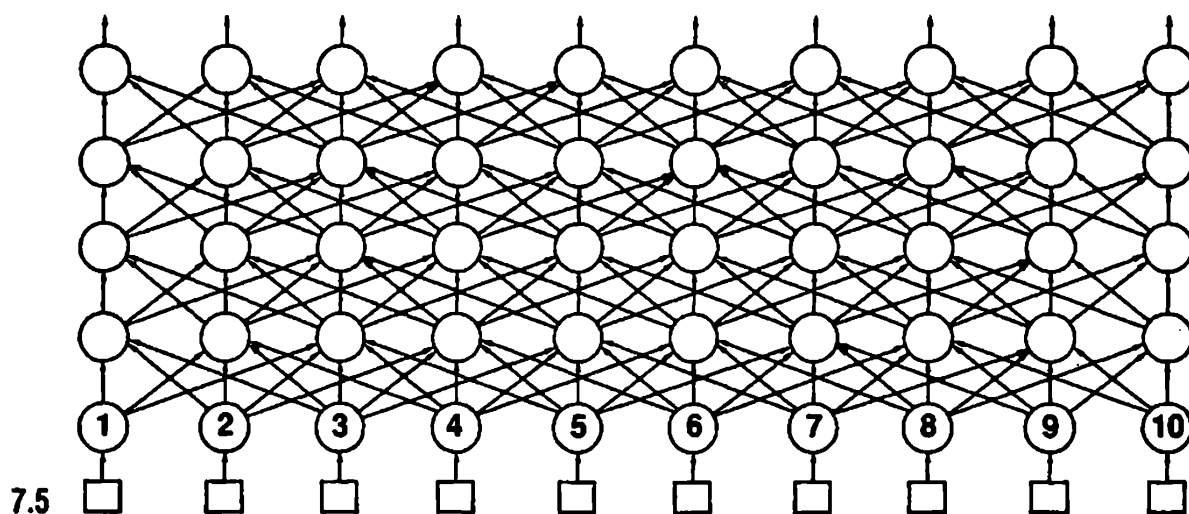
come directory, menu, schermi, shell

connessi da *percorsi* virtuali

come «ESCI e vai al DOS» o *entrare* (ENTER) nel menu STAMPA dal menu PRINCIPALE

e dalla possibilità di eseguire *operazioni* virtuali complesse ed interessanti

come ricercare una parola in un file o ingrandire una «finestra» tracciata sullo schermo.

Poiché qualsiasi macchina calcolatrice può essere imitata da una macchina virtuale su una macchina di von Neumann, ne consegue che se il cervello è una macchina che svolge un'imponente elaborazione in parallelo, anch'esso può essere imitato perfettamente da una macchina di von Neumann. E fin dagli albori dell'era dei calcolatori, i teorici hanno usato questo potere camaleontico delle macchine di von Neumann per creare architetture *virtuali* in parallelo che dovrebbero imitare strutture simili a quelle cerebrali.[17] Come puoi trasformare una macchina che fa una cosa alla volta in una macchina che fa tante cose contemporaneamente? Tramite un processo che assomiglia al lavoro a maglia. Si supponga di dover simulare

7.5

un'elaborazione parallela a dieci canali. Dapprincipio, la macchina di von Neumann riceve l'istruzione di eseguire l'operazione prevista al primo nodo del primo canale (il nodo 1 del disegno) e di salvare il risultato in una memoria «di transito», poi di andare al nodo 2, e così via, fino a che tutti e dieci i nodi del primo strato non siano fatti avanzare di un passo. Poi la macchina di von Neumann affronta gli effetti che questi risultati del primo strato hanno sullo strato successivo di nodi, traendo, uno alla volta, i risultati precedentemente calcolati dalla memoria di transito e inserendoli come dati in ingresso nel secondo strato. Così procedendo laboriosamente e andando avanti e indietro come una vecchietta che fa la calza, baratta il tempo per lo spazio. Per simulare una macchina virtuale a dieci canali occorrerà *un tempo almeno* dieci volte più lungo di quello che occorre per simulare una macchina a un canale e per simulare una macchina a un milione di canali (come il cervello) occorrerà un tempo almeno un milione di volte più lungo. La dimostrazione di Turing non dice nulla sulla velocità con la quale l'imitazione verrà realizzata, e, per alcune architetture, perfino l'abbagliante velocità dei moderni calcolatori numerici viene sopraffatta dal compito. Ecco perché i ricercatori di IA interessati a esplorare le potenzialità delle architetture parallele si stanno oggi volgendo verso *reali* macchine parallele – manufatti che possono essere più giustamente chiamati «giganteschi cervelli elettronici» – su cui effettuare le loro simulazioni. Ma, in linea di principio, qualunque macchina parallela può essere imitata perfettamente – sebbene inefficacemente – come macchina virtuale su una macchina seriale di von Neumann.[18]

Ora siamo pronti per rovesciare questa idea diffusa. Proprio come puoi simulare un cervello parallelo su una macchina seriale di von Neumann, così puoi anche, in linea di principio, simulare (qualcosa che assomiglia a) una macchina di von Neumann su un hardware parallelo, e questo è proprio quello che sto suggerendo: le menti umane coscienti sono delle macchine virtuali più o meno seriali implementate – inefficacemente – sull'hardware parallelo che l'evoluzione ci ha fornito.

Cosa verrà considerato come il «programma» quando si parla di una macchina virtuale che gira sull'hardware parallelo del cervello? Ciò che si deve tener presente è l'enorme plasticità regolabile che può dar luogo a una miriade di microabitudini differenti e quindi dar luogo a differenti macroabitudini. Nel caso della macchina di von Neumann, ciò viene realizzato tramite centinaia di migliaia di zero e uno (bit), divisi in

«parole» di 8, 16, 32 o 64 bit, a seconda della macchina. Le parole sono immagazzinate separatamente nei registri della memoria e vengono trasferite una alla volta nel registro delle istruzioni. Nel caso della macchina parallela ciò viene realizzato, possiamo supporre, tramite migliaia o milioni o miliardi di connessioni che vengono stabilite tra i neuroni e che danno all'hardware sottostante, tramite la loro unione concertata, un nuovo insieme di macroabitudini, un nuovo insieme di regole condizionali del comportamento.

E come vengono installati nel calcolatore cerebrale questi programmi di milioni di connessioni neuronali? In una macchina di von Neumann, basta «caricare» il programma da un dischetto alla memoria centrale e il calcolatore ottiene così un insieme istantaneo di nuove abitudini; con i cervelli, serve l'*allenamento*, comprendente soprattutto l'autostimolazione ripetitiva del tipo abbozzato nel paragrafo 5. Questa, naturalmente, è un'importante discrepanza nella nostra analogia. L'unità centrale di elaborazione o CPU (*Central Processing Unit*) di una macchina di von Neumann risponde in modo rigido alle stringhe di bit che compongono le sue parole e le tratta come istruzioni in un *linguaggio macchina* del tutto specifico e determinato. Questo è ciò che definisce il *calcolatore numerico con programma immagazzinato*, e il cervello umano non è una cosa del genere. Sebbene sia probabilmente vero che ogni particolare connessione stabilita tra i neuroni del cervello ha un effetto determinato sul comportamento risultante della rete circostante, non c'è nessuna ragione di credere che due cervelli differenti abbiano lo «stesso sistema» di interconnessioni, e così, quasi certamente, non c'è nulla di seppur lontanamente analogo al ben definito linguaggio macchina che condividono, per esempio, tutti i calcolatori IBM o IBM-compatibili. Così se due o più cervelli «si passano il software», ciò non avviene in virtù di un processo semplice e diretto analogo alla copia di un programma in linguaggio macchina dall'una all'altra memoria. (Inoltre, naturalmente, la plasticità che in qualche modo favorisce la memoria nel cervello non è isolata come un magazzino passivo; la divisione del lavoro tra memoria e CPU è un artefatto che non ha un analogo nel cervello, ma su questo ritorneremo nel capitolo 9.)

Poiché ci sono queste importanti – e spesso trascurate – discrepanze che indeboliscono l'analogia, perché insisto a paragonare la coscienza umana al software? Perché, come spero di far vedere, alcune caratteristiche importanti e estremamente

enigmatiche della coscienza ricevono un'illuminante spiegazione ammettendo l'ipotesi che la coscienza umana (1) sia un'innovazione troppo recente per essere cablata rigidamente nel meccanismo innato, (2) sia in larga parte un prodotto dell'evoluzione culturale che viene trasmessa ai cervelli nel primo stadio dei loro allenamenti, e (3) che la riuscita della sua installazione sia determinata da una miriade di microaggiustamenti nella plasticità del cervello, il che significa che le sue caratteristiche funzionalmente importanti sono molto probabilmente invisibili a uno scrutinio neuroanatomico malgrado l'estrema rilevanza degli effetti. Proprio come nessun informatico si sognerebbe di stabilire i rispettivi punti di forza e di debolezza di programmi come WordStar e WordPerfect partendo dalle informazioni sulle differenze di voltaggio nella memoria, così nessuno scienziato cognitivista dovrebbe aspettarsi di arrivare a qualcosa di sensato sulla coscienza umana basandosi semplicemente sulla neuroanatomia. Inoltre, (4) l'idea dell'*illusione dell'utente* di una macchina virtuale è stuzzicante e suggestiva: se la coscienza è una macchina virtuale, chi è l'utente per il quale si forma l'illusione? Ammetto che questo sembra sospetto, come se fossimo inesorabilmente risospinti verso un interno Sé Cartesiano, che siede alla consolle della stazione di lavoro corticale e reagisce alla illusione dell'utente del software che sta girando, ma ci sono, come vedremo, dei modi per sfuggire a questa terribile conclusione.

Supponiamo allora per un momento che nella memosfera esista una versione più o meno priva di errori («*debugged*») di questa macchina virtuale del flusso della coscienza – la macchina joyceana. Poiché, come abbiamo visto, non c'è un linguaggio macchina comune tra cervelli, i metodi di trasmissione che garantirebbero una macchina virtuale piuttosto uniforme e operante attraverso l'intera cultura devono essere sociali, molto sensibili al contesto e in un certo grado auto-organizzati e autocorrettivi. Far in modo che due calcolatori differenti – per es., un Macintosh e un PC-IBM – «si parlino tra loro» è un problema ingegneristico intricato e meticoloso la cui soluzione dipende dal grado di precisione delle informazioni che si hanno sui meccanismi interni dei due sistemi. Se gli esseri umani possono in una certa misura «scambiarsi il software» senza che nessuno abbia una tale conoscenza, ciò è dovuto forse al fatto che i sistemi scambiati hanno un alto grado di instabilità e di tolleranza nel formato. Ci sono vari sistemi per passarsi il software: apprendere per imitazione, apprendere come risultato

di un «rinforzo» (o deliberatamente imposto dall'insegnante – ricompensa, incoraggiamento, disapprovazione, minaccia – o trasmesso sottilmente e inconsciamente nel corso degli incontri comunicativi), e apprendere come risultato di un'istruzione esplicita nel linguaggio naturale che è già stato appreso attraverso i due metodi precedenti. (Pensa, per esempio, al tipo di abitudine che si acquisirebbe dicendo frequentemente a un apprendista «Dimmi cosa stai facendo» e «Dimmi perché stai facendo questo». Ora pensa all'apprendista che acquista l'abitudine di indirizzare le stesse richieste a se stesso.)

In realtà, non solo il linguaggio parlato ma anche quello scritto gioca, credo, un ruolo fondamentale nello sviluppo e nell'elaborazione delle macchine virtuali che la maggior parte di noi fa girare la maggior parte del tempo nei nostri cervelli. Proprio come la ruota è un prezioso ritrovato tecnologico che dipende in qualche misura da raggi o strade lastricate o altre superfici artificialmente spianate per la sua utilizzazione, così la macchina virtuale di cui sto parlando può esistere solo in un ambiente in cui ci sia non solo il linguaggio e l'interazione sociale, ma anche lo scrivere e il disegnare, semplicemente perché le esigenze riguardo alla memoria e al riconoscimento di configurazioni necessarie per la sua implementazione richiedono che il cervello «scarichi» alcune delle sue memorie in «memorie di transito» dell'ambiente. (Si noti che ciò implica che la «mentalità preletterata» possa benissimo implicare una classe significativamente differente di architetture virtuali da quelle che si incontrano in società alfabetizzate.)

Prova a sommare due numeri di dieci cifre a mente, senza usare carta e matita *o* ripetere ad alta voce i numeri. Prova a raffigurarti, senza fare un disegno, come realizzare un'intersezione a quadrifoglio di tre autostrade in modo che si possa andare da qualunque direzione di qualunque autostrada a qualunque direzione di qualunque autostrada senza passare per la terza autostrada. Questi sono i tipi di problemi che gli esseri umani risolvono facilmente con l'aiuto di supporti di memoria esterni e l'uso dei loro preesistenti analizzatori (chiamati *occhi e orecchi*) con circuiti di riconoscimento di configurazioni che sono altamente sviluppati e rigidamente cablati. (Vedi Rumelhart, capitolo 14, in McClelland e Rumelhart, 1986, per alcune preziose osservazioni sull'argomento.)

Nel corso del primo sviluppo infantile, noi installiamo nel nostro cervello un insieme, ben organizzato e parzialmente già sperimentato, di abitudini della mente, come le chiama il

politologo Howard Margolis (1987). Nel capitolo 9 osserveremo più da vicino i verosimili dettagli di questa architettura, ma appare chiaro fin da ora che la struttura generale del nuovo insieme di regole consiste di un *concatenamento seriale*, in cui (più o meno) nello stesso «luogo» accade prima una «cosa» e poi un'altra «cosa». In questo flusso di eventi si collocano innumerevoli abitudini apprese, di cui «il parlare a se stessi» è il primo esempio.

Poiché questa nuova macchina creata in noi è un complesso di memi altamente replicato, potremmo chiederci a cosa sia dovuto il suo successo replicativo. Dovremmo, naturalmente, tener presente che non è necessario che ciò avvenga per il *bene* di qualcuno – forse solo per quello della replicazione. *Potrebbe* essere un virus informatico, un parassita che si inserisce facilmente nei cervelli umani senza dare effettivamente nessun vantaggio competitivo agli esseri di cui infesta i cervelli. O più plausibilmente, *alcune caratteristiche* della macchina potrebbero essere dei parassiti, che esistono solo perché possono esistere e perché non è possibile – o non vale la pena – sbarazzarsi di loro. William James pensava che sarebbe assurdo supporre che la cosa più sorprendente che conosciamo nell'universo – la coscienza – sia un mero epifenomeno che non gioca nessun ruolo essenziale nel funzionamento del nostro cervello, ma, per quanto inverosimile ciò possa sembrare, non è del tutto fuori discussione e quindi non è realmente assurdo. Esistono innumerevoli prove sui benefici che la coscienza ci fornisce, quindi possiamo senza dubbio tranquillizzarci sulle sue varie *raisons d'être*, ma finiremmo per interpretare erroneamente tali prove se pensassimo che si rimane nel mistero fino a che ogni singola caratteristica non abbia – o abbia avuto – una funzione (dal *nostro* punto di vista di «utenti» della coscienza) (Harnad, 1982). È possibile che alcuni fatti bruti siano del tutto privi di qualsiasi giustificazione funzionale. Alcune caratteristiche della coscienza potrebbero essere semplicemente dei memi egoisti.

Prendendo in considerazione il lato positivo, comunque, quali problemi questa macchina sembra ben progettata a risolvere? Lo psicologo Julian Jaynes (1976) ha sostenuto persuasivamente che le sue capacità di auto-esortazione e auto-rammemorazione costituiscono i prerequisiti di elaborati progetti a lungo termine senza i quali l'agricoltura, l'edilizia e altre attività civilizzate e civilizzanti non potrebbero essere organizzate. Sembra essere utile anche per quel tipo di auto-controllo che impedisce ad un sistema avariato di essere travolto dalle pro-

prie avarie, un tema questo sviluppato nel campo dell'Intelligenza Artificiale da Douglas Hofstadter (1985). Lo psicologo Nicholas Humphrey, inoltre, ha proposto che essa possa fornire un mezzo per sfruttare ciò che potremmo chiamare la simulazione sociale – l'uso dell'introspezione per guidare le proprie impressioni su ciò che gli altri stanno pensando e sentendo.

Alla base di questi talenti avanzati e specializzati c'è la capacità fondamentale di risolvere il meta-problema del «a cosa pensare dopo». Abbiamo già visto in questo capitolo che quando un organismo fronteggia una crisi (o semplicemente un problema nuovo e difficile), può avere al suo interno delle risorse che sarebbero preziose in quelle circostanze *se solo potesse trovarle e utilizzarle in tempo!* Le risposte di orientamento, come Odmar Neumann ha ipotizzato, hanno l'effetto positivo di mettere quasi tutti al lavoro nello stesso tempo, ma elevando in questo modo il livello globale di vigilanza si contribuisce, come abbiamo visto, sia alla soluzione del problema sia ad un'ulteriore complicazione. Non è infatti di nessun aiuto se, successivamente, il cervello non riesce ad ottenere una qualche attività coerente da tutti questi volontari. Il problema che aveva generato come sua soluzione parziale le risposte di orientamento era quello della necessità di un accesso reciproco e globale tra un gruppo di specialisti abituati a preoccuparsi solo dei propri affari. Anche se, grazie ad una sottostante architettura in stile pandemonio, il caos presto si quieta, lasciando temporaneamente solo uno specialista in carica (e, forse, meglio informato grazie alla competizione che ha vinto), ci sono ovviamente tanti modi di risolvere male questi conflitti quanto di risolverli bene. Nulla garantisce che lo specialista più efficiente politicamente sia anche «l'uomo giusto per il problema».

Platone ha visto chiaramente il problema duemila anni fa e ha trovato una mirabile metafora per descriverlo.

Vedi ora se anche della conoscenza è possibile dire così, che uno, pur possedendola, non l'abbia; bensì la possegga come diremmo di uno il quale, presi alla caccia certi uccelli selvatici, colombi o altro, se li allevi in casa dentro una colombaia che s'è costruita: del quale solo in un certo senso potremmo dire che ha sempre codesti uccelli, in quanto già li possiede. O no? [...] Se non che, in altro senso, potremmo anche dire che non ne ha nessuno, ma solo ha sopra di essi, dal momento che se li soggettò collocandoli in un suo recinto

domestico, la potestà di prendere e tenere, quando voglia, quel qualunque uccello che di volta in volta brami cacciare, e poi di nuovo rilasciarlo: e questo lo può fare tutte le volte che gli piace [...] immaginiamo in ciascuna anima una specie di gabbia ripiena di uccelli di ogni sorta, dei quali parte vadano a schiere, separatamente dagli altri, parte a piccoli gruppi, e alcuni solitari, in mezzo a questi o a quelli, svolazzando dove loro capita. (*Teeteto*, 197-198a)

Ciò che Platone ha capito è che avere semplicemente gli uccelli non basta; ciò che è importante e difficile è imparare a far avvicinare l'uccello giusto quando lo chiami. Egli ha sostenuto successivamente che, *ragionando*, miglioriamo la nostra abilità a prendere gli uccelli giusti al momento giusto. Imparare a ragionare, in effetti, equivale a imparare delle strategie di recupero delle conoscenze.[19] Ecco dove le abitudini della mente entrano in gioco. Abbiamo già visto in linee molto generali come tali abitudini generali della mente come «parlare a se stessi» o «raffigurarsi» *potrebbero* pungolare i giusti pezzetti di informazione alla superficie (la superficie di che? – ne riparleremo nel capitolo 10). Ma abitudini della mente più specifiche, miglioramenti ed approfondimenti di modi particolari di parlare a se stessi, possono migliorare ulteriormente le tue possibilità.

Il filosofo Gilbert Ryle, nel suo libro pubblicato postumo *On Thinking* (1979), decise che il pensare lento e difficile, quel ponderare in cui la famosa statua di Rodin *Il pensatore* sembra chiaramente essere immersa, deve spesso essere effettivamente un parlare a se stesso. Sorpresa, sorpresa! Non è ovvio che questo è ciò che facciamo quando pensiamo? Sì e no. È ovvio che (spesso) sembra proprio quello che facciamo: spesso possiamo perfino dirci l'un l'altro le parole che usiamo nei nostri silenziosi soliloqui. Ma ciò che è tutt'altro che ovvio è perché parlare a se stessi è utile.

Che cosa sta facendo *Il pensatore* nel suo interno apparentemente cartesiano? O, per sembrare scientifici, a cosa assomigliano i processi mentali che si stanno svolgendo in quella *camera obscura* cartesiana? [...] Notoriamente alcune nostre ponderazioni, ma non tutte, terminano con la soluzione dei nostri problemi; eravamo confusi, ma alla fine abbiamo raggiunto la chiarezza. Ma se talvolta abbiamo successo, perché non sempre? Se tardivamente, perché non prontamente? Se difficilmente, perché non facilmente? Perché mai funziona in effetti? Come è possibile che funzioni? (Ryle, 1979, p. 65)

Le abitudini della mente sono state progettate nei millenni per *modellare* dei percorsi praticabili di esplorazione. Come Margolis nota,

> persino un essere umano oggi (e quindi, *a fortiori*, un remoto antenato dei contemporanei esseri umani) non può facilmente o ordinariamente mantenere un'attenzione ininterrotta su di un singolo problema per più di poche decine di secondi. Eppure lavoriamo su problemi che richiedono un tempo enormemente superiore. Ciò richiede (come possiamo notare osservando noi stessi) fasi di rimuginamento seguite da fasi di ricapitolazione, che servono a descrivere a noi stessi ciò che sembra essere avvenuto durante il rimuginamento, che ha condotto a un qualche risultato intermedio. Ciò ha una funzione ovvia: ripetendo questi risultati provvisori [...] noi li affidiamo alla memoria, giacché i contenuti immediati del flusso della coscienza vengono persi molto rapidamente se non vengono ripetuti [...] Dato il linguaggio, possiamo descrivere a noi stessi che cosa sembrava avvenire durante il rimuginamento che condusse al giudizio, produrre una versione ripetibile del processo che ci ha condotto al giudizio e affidarla alla memoria a lungo termine grazie al fatto di ripeterla. (Margolis, 1987, p. 60)

È qui, nelle abitudini individuali di autostimolazione, che dobbiamo cercare i *kludge*, il termine con cui gli *hacker*, i patiti dell'informatica, indicano quegli espedienti *ad hoc* che di solito vengono inseriti come rappezzature nel software durante il *debugging* per far funzionare davvero i programmi. [La linguista Barbara Partee una volta criticò un'inelegante rappezzatura in un programma di IA per l'analisi linguistica chiamandola «*odd hack*» (strano pasticcio, bizzarra trovata) – uno dei più meravigliosi giochi di parole o scambi di lettere involontari che io abbia mai incontrato. Madre Natura è piena di bizzarre trovate e dovremo aspettarci di trovarle anche nella adozione personalizzata della macchina virtuale da parte dell'individuo].

Ecco un esempio plausibile: poiché la memoria umana non è ben progettata fin dalla nascita per essere super-affidabile e di rapido accesso (come la RAM necessaria in ogni macchina di von Neumann), quando i progettisti (distribuiti nel tempo e nella cultura) della macchina virtuale neumanniana si trovarono di fronte al problema di abborracciare un sostituto adatto che girasse nel cervello, si sono imbattuti in vari trucchi che potenziano la memoria. I trucchi fondamentali si basano sulla ripetizione, ripetizione e ancora ripetizione, favorita dalle rime e dal ritmo, e da massime facilmente ricordabili. (Le rime e il ritmo sfruttano il grande potere dei sistemi di analisi uditiva

pre-esistenti di riconoscere le configurazioni sonore.) La deliberata e ripetuta giustapposizione di elementi tra i quali si doveva stabilire un legame associativo – cosicché una cosa avrebbe sempre «ricordato» l'altra al cervello – fu potenziata ulteriormente, possiamo supporre, rendendo le associazioni più ricche possibile, rivestendole non solo di caratteristiche visive e uditive, ma sfruttando l'intero corpo. L'aggrottare le ciglia e il tenersi il mento del *Pensatore*, il grattarsi la testa, il mormorare, l'andare su e giù, lo scarabocchiare che noi tutti, a seconda del carattere, più o meno facciamo, potrebbero rivelarsi non semplici prodotti collaterali casuali e inutili del pensiero cosciente, ma collaboratori funzionali (o le vestigia di precedenti e più rozzi collaboratori funzionali) per la laboriosa disciplina che il cervello doveva intraprendere per trasformarsi in una mente matura.

E al posto di un sistematico e preciso «ciclo di prelievo» o «ciclo di istruzione» che porta ogni nuova istruzione al registro delle istruzioni per essere eseguita, dovremmo aspettarci delle «regole» di transizione poco logiche, alquanto imprecise e mal ordinate, dove la forte inclinazione innata del cervello per «la libera associazione» è rifornita di lunghe catene associative per dare delle garanzie abbastanza buone che la giusta sequenza venga sottoposta a prova. (Nel capitolo 9 prenderemo in esame alcuni approfondimenti di quest'idea nell'IA; per approfondimenti con enfasi differenti, vedi Margolis, 1987 e Calvin, 1987, 1989. Vedi anche Dennett, 1991b.) Non dovremmo aspettarci che la maggior parte delle sequenze che prendono corpo siano degli *algoritmi* ben dimostrati, che sicuramente conducono ai risultati cercati, ma solo incursioni guidate non esclusivamente dal caso nella colombaia di Platone.

L'analogia con le macchine virtuali dell'informatica fornisce una prospettiva utile per lo studio del fenomeno della coscienza umana. Originariamente i calcolatori sono stati considerati solo dei divoratori di numeri, ma ora questa loro capacità è stata imbrigliata in migliaia di modi fantasiosi per creare nuove macchine virtuali, come i videogiochi e i *word processor*, in cui la sottostante capacità di divorare numeri è quasi invisibile e in cui i nuovi poteri sembrano quasi magici. Similmente, i nostri cervelli non erano progettati (tranne che per qualche organo periferico molto recente) per l'elaborazione delle parole (il *word processing*), ma ora una porzione molto ampia – forse anche la parte del leone – dell'attività che si svolge nei cervelli umani adulti ha a che fare in qualche modo con l'elaborazione delle

parole: la produzione e la comprensione di discorsi, la ripetizione seriale e il riordinariamento di oggetti linguistici, o meglio dei loro surrogati neurali. E queste attività amplificano e trasformano i poteri dell'hardware sottostante in maniere che («dall'esterno») sembrano magiche.

Ma comunque – sono sicuro che vorrai obiettare – tutto ciò ha poco o nulla a che fare con la coscienza! Dopo tutto, una macchina di von Neumann è completamente inconscia; perché mai una sua implementazione – o qualcosa di simile a essa, una macchina joyceana – dovrebbe essere più conscia? Ho una risposta: la macchina di von Neumann, essendo cablata fin dall'inizio in quel modo, con legami informazionali della massima efficienza, non deve diventare l'oggetto dei suoi propri elaborati sistemi percettivi. Il lavorio della macchina joyceana, d'altra parte, è altrettanto «visibile» e «udibile» a se stessa di qualunque altra cosa nel mondo esterno che essa può percepire – per la semplice ragione che su tale lavorio è concentrata gran parte dello stesso meccanismo percettivo.

Ora ciò *appare* un bel gioco di specchi, lo so. E certamente è controintuitivo, difficile da mandar giù, inizialmente offensivo – proprio come ci aspetteremmo da un'idea che possa aprirsi un varco attraverso secoli di mistero, controversie e confusione. Nei prossimi due capitoli vedremo più da vicino – e scetticamente – il modo in cui questo trucco apparente potrà rivelarsi una parte legittima di una spiegazione della coscienza.

8

COME LE PAROLE COLLABORANO CON NOI

> Il linguaggio, come la coscienza, sorge soltanto dal bisogno, dalla necessità dei rapporti con gli altri.
>
> KARL MARX, 1846

> La coscienza in generale si è sviluppata soltanto sotto la pressione del bisogno di comunicazione.
>
> FRIEDRICH NIETZSCHE, 1882

> Prima che il mio maestro venisse a me, non sapevo di esistere. Vivevo in un mondo che era un non-mondo. Non posso sperare di descrivere adeguatamente quel tempo inconscio, eppur conscio, di vacuità... Poiché non avevo alcun potere intellettuale, non posso paragonare uno stato mentale all'altro.
>
> HELEN KELLER, 1908

1. *Facciamo il punto: «E pluribus unum»?*

Nel capitolo 5 abbiamo esposto l'idea, cattiva ma ostinatamente attraente, del Teatro Cartesiano, in cui uno spettacolo di suoni e luci viene presentato a un pubblico esclusivo e potente, l'Io o il Dirigente Superiore. Anche se abbiamo visto le incoerenze di questa idea, e identificato un'alternativa, il modello delle Molteplici Versioni, il Teatro Cartesiano continuerà a perseguitarci fino a che non avremo ancorato la nostra alternativa alle solide fondamenta della scienza empirica. Questo compito è stato avviato nel capitolo 6, e nel capitolo 7 abbiamo fatto ulteriori progressi. Siamo tornati, letteralmente, ai princìpi primi: i princìpi dell'evoluzione che ci hanno guidato in una ricostruzione ipotetica del processo graduale dello sviluppo progettuale che ha creato il nostro tipo di coscienza. Questo ci ha permesso di gettare un'occhiata ai meccanismi della coscienza dall'interno della scatola nera; da dietro le quinte, potremmo dire, in onore dell'allettante immagine teatrale che stiamo tentando di rovesciare.

Nel nostro cervello c'è un'aggregazione un po' abborracciata di circuiti cerebrali specializzati, che, grazie a svariate abitudini indotte in parte dalla cultura e in parte dall'auto-

esplorazione individuale, lavorano assieme alla produzione più o meno ordinata, più o meno efficiente, più o meno ben progettata di una macchina virtuale, la *macchina joyceana.* Facendo lavorare per una causa comune questi organi specializzati che si sono sviluppati indipendentemente, e dando quindi alla loro unione dei poteri ampiamente potenziati, questa macchina virtuale, questo software del cervello, opera una sorta di miracolo politico interno: crea un comandante *virtuale* dell'equipaggio, senza conferire a nessuno dei poteri dittatoriali a lungo termine. Chi comanda? Prima una coalizione e poi un'altra, alternandosi ordinatamente al potere grazie a buone meta-abitudini che tendono a produrre successioni al trono coerenti e finalizzate piuttosto che caotici e interminabili colpi di Stato.

La saggezza esecutiva che ne risulta è solo uno dei poteri tradizionalmente attribuiti al Sé, ma è molto importante. William James gli ha reso omaggio quando ha preso in giro l'idea di un Neurone Pontificio in qualche parte del cervello. Noi sappiamo che la descrizione del lavoro che spetterebbe a un tale sottosistema Boss è incoerente, ma sappiamo anche che quelle responsabilità e decisioni di controllo devono essere *in qualche modo* parcellizzate nel cervello. Noi *non* siamo navi alla deriva con equipaggi rissosi; riusciamo abbastanza bene non solo a stare alla larga da secche e altri pericoli, ma anche a pianificare campagne, correggere errori tattici, riconoscere vaghi messaggeri di opportunità e controllare enormi progetti che si dispiegano nel corso di mesi o anni. Nei prossimi capitoli osserveremo più da vicino l'architettura di questa macchina virtuale, per poter suffragare – non dimostrare – l'ipotesi che possa effettivamente svolgere queste funzioni esecutive ed altre. Prima di far questo, comunque, dobbiamo esporre e neutralizzare un'altra fonte di mistificazioni: l'illusione di un Autore Centrale (*Central Meaner*).

Uno dei compiti fondamentali del Boss immaginario è quello di controllare le comunicazioni con il mondo esterno. Come abbiamo visto nel capitolo 4, l'eterofenomenologia è resa possibile dalla presupposizione che ci sia qualcuno dentro casa che parla, che intende dire ciò che viene detto, che è l'autore ufficiale delle dichiarazioni rilasciate. Quando ci accingiamo ad interpretare i suoni vocali di un corpo, non pensiamo che siano solo delle strida casuali o delle parole prodotte miracolosamente dallo schiamazzo di un'orgia dietro le quinte, ma li consideriamo gli atti di un singolo agente, di quella (unica) *persona* il cui corpo emette i suoni. Una volta che abbiamo scelto di interpre-

tare, non possiamo far altro che postulare una persona di cui stiamo interpretando gli atti comunicativi. Questo non è del tutto equivalente al postulare un sistema *interno* che sia il Boss del corpo, il Burattinaio che manovra i burattini, ma ciò nonostante questa è l'immagine che naturalmente ci cattura. Questo Boss interno, siamo tentati di supporre, sembra più o meno come il presidente degli Stati Uniti, che può incaricare qualche addetto stampa o altri funzionari di rilasciare le effettive dichiarazioni, ma quando questi parlano, lo fanno per suo conto, eseguono i suoi atti linguistici, per i quali egli è il responsabile e dei quali egli è, ufficialmente, l'autore.

Non esiste in realtà una tale gerarchia nel cervello che presiede alla produzione dei discorsi (o degli scritti). Se vogliamo davvero smantellare il Teatro Cartesiano dobbiamo, tra le altre cose, trovare una spiegazione più realistica della reale fonte (fonti) delle affermazioni, domande e altri atti linguistici che naturalmente attribuiamo alla (singola) persona il cui corpo emette le vocalizzazioni. Dobbiamo capire che cosa succede al proficuo mito dell'eterofenomenologia quando diamo alle complessità della produzione linguistica l'attenzione dovuta.

Abbiamo già visto un'ombra gettata da questo problema. Nel capitolo 4, abbiamo immaginato che il robot Shakey avesse una rudimentale capacità di conversare o almeno di emettere parole in determinate circostanze. Avevamo supposto che Shakey potesse essere progettato in modo da «dirci» come discrimina le scatole dalle piramidi. Shakey potrebbe dire «Analizzo ogni sequenza di 10.000 cifre...» o «Cerco i confini tra il chiaro e lo scuro e li evidenzio con delle linee...» o «Non lo so; alcune cose mi appaiono semplicemente a forma di scatola...». Ognuno di questi «resoconti» differenti scaturirebbe da un livello particolare di accesso tra i vari livelli di accesso che il meccanismo «addetto alle comunicazioni» potrebbe avere al lavorio del meccanismo «di identificazione delle scatole», ma non ci eravamo spinti nei dettagli del modo in cui i vari stati di macchina interni sarebbero agganciati alle esternazioni che producono. Questo era un modello volutamente semplificato dell'effettiva produzione linguistica e aveva solo lo scopo di mettere in luce un punto molto astratto del nostro esperimento di pensiero: se un sistema di emissione di enunciati ha solo un accesso limitato ai suoi stati interni e ha un vocabolario limitato con cui comporre i suoi enunciati, si possono interpretare come veri i suoi «resoconti» solo se ne diamo una lezione in qualche modo

metaforica. Le «immagini» di Shakey fornivano un esempio di come qualcosa che in realtà non era affatto un'immagine potesse essere proprio ciò di cui si stava realmente parlando sotto la maschera di un'immagine.

Una cosa è aprire una possibilità astratta, un'altra è far vedere che questa possibilità ha una versione realistica che si applica a noi. Ciò che Shakey faceva non era un vero resoconto, un vero parlare. A quanto sembra, l'immaginaria verbalizzazione di Shakey sarebbe simile al linguaggio artificioso, «preconfezionato» che i programmatori inseriscono nei programmi che presentano un interfaccia amichevole con l'utente. Stai per formattare un dischetto e il tuo calcolatore ti «chiede» amichevolmente: «Sei sicuro che lo vuoi fare? Tutto il contenuto del disco verrà cancellato! Rispondi S o N». Solo un utente molto ingenuo potrebbe pensare che il calcolatore realmente *intendesse* essere così premuroso.

Mi si permetta di mettere qualche parola in bocca a un critico. Poiché questo critico immaginario ci accompagnerà nelle discussioni e nelle ricerche dei prossimi capitoli, io gli darò un nome. Otto dice:

È stato un trucco a buon mercato chiamare Shakey «egli» piuttosto che «esso»; il problema con Shakey è che non ha una vera interiorità come noi; non si prova nulla ad essere Shakey. Anche se il macchinario che prende i segnali in entrata dal suo «occhio» televisivo e li utilizza per identificare le scatole fosse fortemente analogo a quello del *nostro* sistema visivo (e non è così), e anche se il macchinario che controlla la sua produzione di stringhe di parole inglesi fosse fortemente analogo a quello del *nostro* sistema linguistico che controlla la produzione di stringhe di parole inglesi (e non è così), ci sarebbe *ancora* qualche cosa che manca: l'Uomo-Intermedio che sta in ognuno di noi e i cui giudizi vengono espressi quando diciamo cosa accade in noi. Il problema con Shakey è che i suoi input ed output sono collegati tra loro nel modo sbagliato – un modo che elimina l'osservatore (colui che soffre e gioisce) che deve trovarsi in qualche punto tra il suo input visivo e il suo output verbale, così che ci sia qualcuno lì che *intenda dire* le parole di Shakey quando vengono «dette».

Quando io parlo [seguita Otto] intendo dire ciò che dico. La mia vita cosciente è privata, ma posso decidere di divulgarne alcuni aspetti; posso decidere di dirti varie cose sulla mia esperienza attuale o passata. Quando lo faccio, formulo delle

frasi che confeziono attentamente sul materiale che voglio riferire. Posso fare la spola tra l'esperienza e il resoconto che sto per formulare, confrontando le parole con l'esperienza per essere sicuro di aver trovato *les mots justes*. Questo vino ha una punta di *pompelmo* nel suo gusto o mi fa pensare di più alle *fragole*? Sarebbe più giusto dire che la nota più alta era *più forte* o in realtà sembrava soltanto *più chiara* o *meglio definita*? Mi rivolgo alla mia esperienza cosciente e arrivo ad un giudizio sulle parole che meglio renderebbero giustizia alle sue peculiarità. Quando sono soddisfatto, quando ritengo di aver formulato un resoconto accurato, lo esprimo. Grazie al mio resoconto introspettivo, tu puoi venir a sapere alcune caratteristiche della mia esperienza cosciente.

Come eterofenomenologi, dobbiamo dividere questo testo in due parti. Da una parte mettiamo le affermazioni sul modo in cui l'esperienza del parlare appare ad Otto. Queste sono inviolabili; così *è* come l'esperienza sembra ad Otto, e va considerato un dato che aspetta una spiegazione. Dall'altra parte mettiamo le tesi teoretiche (sono le conclusioni di un argomento tacito?) di Otto su ciò che questo evidenzia riguardo a quello che avviene in lui – e in che cosa differisce da quello che avveniva, per esempio, in Shakey. Queste non hanno uno statuto speciale, ma le tratteremo con il rispetto dovuto a tutte le affermazioni sensate.

Per quanto mi riguarda, invece, insisto nel dire che l'Uomo-Intermedio, l'Osservatore Interno del Teatro Cartesiano deve essere eliminato, non trovato, ma non possiamo *semplicemente* buttarlo via. Se non c'è un Autore Centrale, da dove viene il significato? Dobbiamo rimpiazzarlo con una spiegazione plausibile di come un'espressione che si intende dire – un *vero* resoconto, senza virgolette – possa essere composta senza il bisogno dell'*imprimatur* di un solitario Autore Centrale. Questo è il compito centrale di questo capitolo.

2. *Burocrazia contro pandemonio*

Uno degli scheletri nell'armadio della linguistica contemporanea è il fatto di aver profuso attenzione all'ascolto ma di aver ampiamente ignorato il parlare, che, si potrebbe dire, è circa la metà del linguaggio, e la metà più importante. Sebbene ci siano molte teorie dettagliate e molti modelli della *percezione* del

linguaggio e della *comprensione* di una frase ascoltata (la strada che dalla fonologia, attraverso la sintassi, va alla semantica e alla pragmatica), nessuno – né Noam Chomsky né alcuno dei suoi avversari o seguaci – ha mai avuto qualcosa di particolarmente importante (giusto o sbagliato) da dire sui sistemi di *produzione* linguistica. È come se tutte le teorie artistiche fossero teorie sulla *fruizione* artistica senza mai una parola sul momento creativo degli artisti – come se tutta l'arte consistesse di *objets trouvés* apprezzati da galleristi e collezionisti.

Non è difficile capire perché è così. Le frasi *sono* oggetti facilmente rinvenibili con cui iniziare un processo. In realtà è piuttosto chiaro quale sia il materiale grezzo o in ingresso nei sistemi di percezione e comprensione: onde di varie forme nell'aria o stringhe di segni su varie superfici piane. E sebbene il dibattito su che cosa siano i prodotti finali del processo di comprensione sia ancora avvolto in una nebbia profonda, almeno questo profondo disaccordo viene alla fine del processo in esame, non all'inizio. Una corsa con una chiara linea di partenza può almeno essere razionalmente iniziata, anche se nessuno è del tutto sicuro dove andrà a finire. Il «prodotto» o «output» della comprensione del discorso è una *decodifica* o *traduzione* dell'input in una nuova rappresentazione – un enunciato in mentalese, forse, o una fotografia nella testa – o è un insieme di *strutture profonde* o un'entità ancora non immaginata? I linguisti possono decidere di posporre una risposta a questo problemaccio mentre lavorano sulle parti più periferiche del processo.

Con la produzione del discorso, d'altra parte, poiché nessuno finora ha elaborato una descrizione chiara e comunemente accettata di ciò che dà inizio al processo che termina alla fine in un'espressione linguistica completamente sviluppata, è difficile perfino cominciare a formulare una teoria. Difficile, ma non impossibile. Ci sono stati, piuttosto recentemente, dei buoni lavori sulla produzione linguistica, che sono stati confrontati e discussi dallo psicolinguista olandese Pim Levelt nel suo *Speaking* (1989). Lavorando a ritroso dall'output o lavorando dal centro verso le due estremità, possiamo riuscire a gettare uno sguardo piuttosto istruttivo sul macchinario che progetta le nostre frasi e le esprime.

Il discorso non è prodotto «a blocchi», progettando ed eseguendo una parola alla volta. L'esistenza di almeno una limitata capacità di anticipazione nel sistema è rivelata dal modo in cui l'accento tonico viene distribuito in un'espressione. Ecco un esempio inglese tratto dalla discussione di Levelt

(1989). Si noti come la parola «Tennessee» sia accentata in due modi differenti nella frase:

I drove from Nashville, TennesSEE, to the TENnessee border.

(«Ho guidato da Nashville, Tennessee, al confine del Tennessee.»); in italiano tale capacità di anticipazione è molto evidente nel caso dell'elisione; nell'esempio seguente elidiamo la *o* di «quello» a seconda del contesto:

MAESTRO: Chi ti ha dato uno schiaffo?
ALUNNO: È stato quello!
MAESTRO: Quell'alunno è un poco di buono.

Quando il maestro parla per la seconda volta, deve adattare la sua pronuncia di «quello» alla parola che segue. Se avesse detto:

Quello studente è un poco di buono

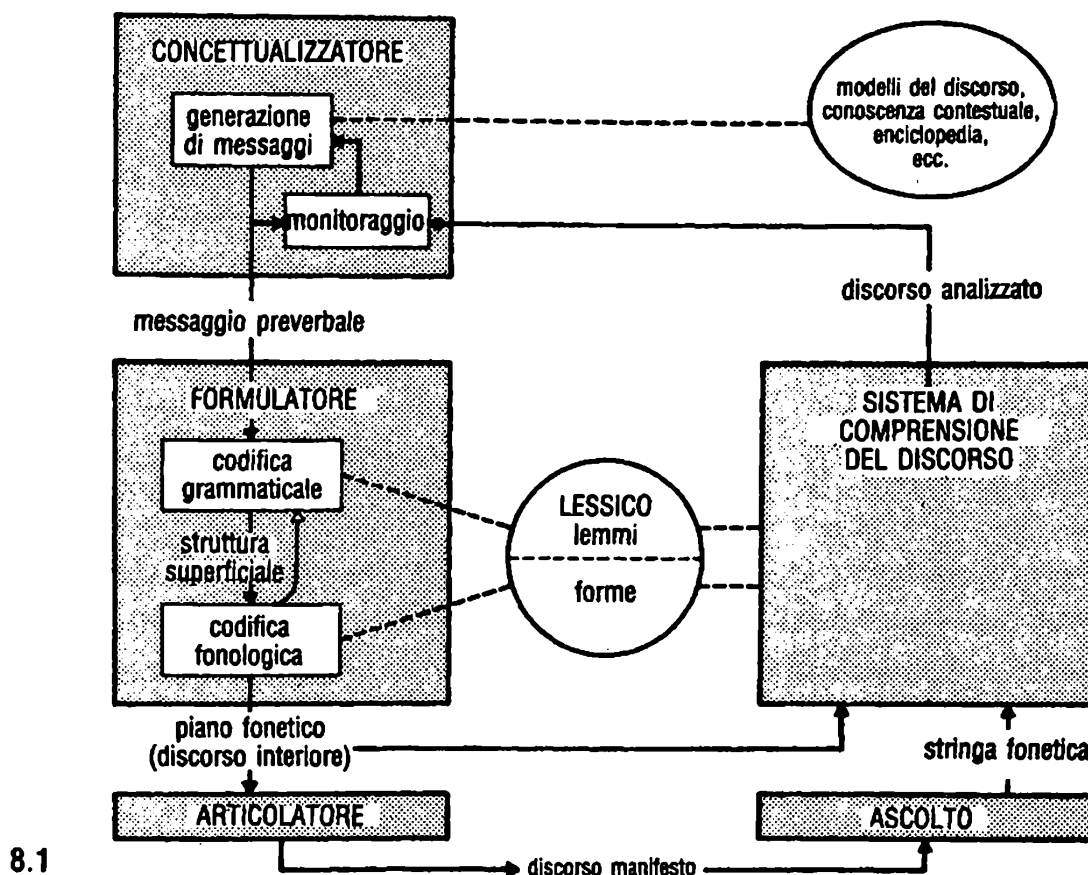
non avrebbe dovuto elidere la «o» finale. Si noti che è più facile accorgersi di questo fenomeno in italiano che in inglese, dal momento che è evidenziato anche da una variazione ortografica.

Scambi di lettere o altri errori linguistici mostrano in modo piuttosto chiaro come le distinzioni lessicali e grammaticali vengono osservate (e non osservate) nel corso della progettazione di una frase da pronunciare. È più facile che la gente dica «lava fessa» quando intenda dire «fava lessa» piuttosto che «fittura presca» quando intende «pittura fresca». C'è una propensione a favore delle parole reali (familiari) rispetto a quelle solamente pronunciabili (possibili ma non reali) anche quando facciamo un *lapsus linguae*. Alcuni errori la dicono lunga sul modo in cui il meccanismo di selezione delle parole deve operare: «Il tuo comportamento è crudale [crudele/brutale]» e «L'ho appena messo in forno a una velocità molto bassa». E pensa alla trasposizione implicata nella produzione di un errore come «Tu canzoni l'amore» per «Tu ami le canzoni».

Grazie a ingegnosi esperimenti che provocano tali errori, e a complicate analisi su ciò che accade o non accade quando la gente parla, sono stati fatti vari progressi nella costruzione di modelli di meccanismi altamente organizzati che eseguono la definitiva articolazione di un messaggio una volta che *sia stato*

deciso di diffondere nel mondo esterno un particolare messaggio. Ma chi o che cosa mette in moto questo meccanismo? Un errore nel discorso è un errore solo perché è differente da ciò che il *parlante intendeva dire*. Quale supervisore *assegna il compito* rispetto al quale errori come gli esempi precedenti vengono giudicati?

Chi, se non l'Autore Centrale? Levelt ci fornisce uno schizzo, un «disegno costruttivo per un parlante»:

8.1

Nell'angolo superiore sinistro un funzionario che ha una somiglianza sospetta con l'Autore Centrale fa la sua comparsa sotto il nome di Concettualizzatore, è armato di un'enorme conoscenza del mondo, di piani e *intenzioni comunicative*, ed è in grado di «generare messaggi». Levelt avverte i suoi lettori che il Concettualizzatore «è una reificazione che abbisogna di ulteriori spiegazioni» (p. 9), ma lo postula in ogni caso, poiché sembra che non possa davvero mettere in moto il processo senza un tale Boss non analizzato che impartisca gli ordini al resto della squadra.

Come funziona? Il problema in questione apparirà con chiarezza se iniziamo con una caricatura. Il Concettualizzatore decide di eseguire un atto linguistico, come insultare il suo interlocutore commentando malevolmente la grandezza dei suoi piedi. Così invia un ordine alla burocrazia sotto di sé,

l'Ufficio per le Pubbliche Relazioni (il Formulatore di Levelt): «Dite a questo cafone che i suoi piedi sono troppo grandi!». Gli addetti alle PR si incaricano del compito. Trovano le parole appropriate: il pronome possessivo singolare di seconda persona, *tuoi*; una buona parola per piedi, come *piedi*; la giusta forma plurale del verbo essere, e cioè *sono*; l'avverbio e l'aggettivo appropriati: *troppo grandi*. Combinano abilmente il tutto, con il giusto tono di voce insultante, ed eseguono:

«I tuoi piedi sono troppo grandi!».

Ma aspetta un attimo. Non è troppo facile? Quando il Concettualizzatore dà l'ordine (ciò che Levelt chiama il *messaggio preverbale*), se lo dà in italiano, come la mia caricatura suggerisce, è lui che ha fatto tutta la parte difficile del lavoro, lasciando al resto della squadra ben poco da fare, tranne qualche banale aggiustamento. Allora il messaggio preverbale viene dato in qualche altro sistema rappresentazionale o linguaggio? Quale che sia, deve essere in grado di fornire alla squadra di produzione la descrizione essenziale dell'oggetto che devono comporre ed emettere, e naturalmente deve essere formulata in termini che *essi* possono «capire» – non in italiano ma in qualche versione di cervellese o mentalese. Dovrà essere una sorta di linguaggio del pensiero, sostiene Levelt, ma forse in un linguaggio del pensiero che viene usato solo per ordinare atti linguistici e non per tutte le attività cognitive. La squadra riceve il messaggio preverbale, un dettagliato ordine in mentalese di pronunciare una frase in italiano, e lo esegue. Questo dà ai sottoposti un po' più di lavoro da svolgere, ma mette in ombra il pericolo di un regresso all'infinito. Come fa il Concettualizzatore a immaginare quali parole mentalesi deve usare per dare l'ordine? Sarebbe bene che non ci fosse un duplicato rimpicciolito dell'intero disegno costruttivo di Levelt nel blocco *generazione messaggi* del Concettualizzatore (e così via, *ad infinitum*). E certamente nessuno ha detto al Concettualizzatore che cosa dire; è l'Autore Centrale, dopo tutto, la *fonte* del significato.

Come si sviluppa allora il significato di una frase? Si consideri allora la seguente gerarchia nidificata di comandi, che va da una strategia globale attraverso una tattica dettagliata alle operazioni elementari:

1. Mettiti sull'offensiva!
2. Cerca di essere cattivo con lui, senza esagerare!

3. Insultalo!
4. Denigra qualche aspetto del suo corpo!
5. Digli che i suoi piedi sono troppo grandi!
6. Dichiara: «I tuoi piedi sono troppo grandi»!
7. Pronuncia: «i 'twɔi̯ 'pjɛdi 'sono 'trɔppo 'grandi».

Sicuramente si deve verificare una qualche approssimazione progressiva all'atto finale simile a questa. Il parlare umano è un'attività finalizzata; ci sono obiettivi e strumenti, e noi in qualche modo facciamo un discreto lavoro vagliando le varie opzioni. Avremmo potuto dargli uno spintone invece di insultarlo o rimpicciolire la sua intelligenza invece di ingrandire i suoi piedi o dire, citando il pianista Fats Waller, «Le tue estremità podali sono detestabili!».

Ma questa approssimazione progressiva viene tutta realizzata tramite una gerarchia burocratica di comandanti che danno ordini ai subordinati? In questa cascata di comandi sembrano esserci molte decisioni che vengono prese – «momenti» in cui le opzioni vengono «scelte» a discapito delle loro rivali – e questo invita a pensare ad un modello in cui c'è una delega di responsabilità per i dettagli minori, in cui gli agenti subordinati con le loro intenzioni apprezzano le ragioni delle varie scelte che fanno. (Se non dovessero affatto capire perché stanno facendo quello che stanno facendo, non sarebbero realmente degli agenti, ma soltanto dei funzionari passivi che si limitano a mettere timbri e lasciano che qualunque cosa passi sulle loro scrivanie abbia autorità su di loro.)

Il disegno costruttivo di Levelt mostra le vestigia di una delle sue fonti: l'architettura di von Neumann che si ispirava alle riflessioni di Turing sul suo flusso di coscienza e che ha ispirato a sua volta molti modelli nella scienza cognitiva. Nel capitolo 7, ho cercato di vincere le resistenze all'idea che la coscienza umana *assomigli* ad una macchina di von Neumann, un elaboratore seriale con una successione di contenuti definiti che si alternano turbinosamente nella strozzatura dell'accumulatore. Ora devo agire sui freni e mettere l'accento su alcuni aspetti dell'architettura funzionale della coscienza umana che *non* sono simili a quella di una macchina di von Neumann. Se confrontiamo il disegno costruttivo di Levelt col modo in cui le macchine di von Neumann usualmente emettono le parole, possiamo accorgerci che il modello di Levelt si attiene troppo alla sua fonte ispiratrice.

Quando una macchina di von Neumann dice quello che sta

scritto nel suo cuore, esterna il contenuto della sua unica officina centrale, l'accumulatore, che ad ogni istante ha un contenuto integralmente specificato nel linguaggio fisso dell'aritmetica binaria. Il rudimentale «messaggio preverbale» di una macchina di von Neumann assomiglia a questo: 10110101 00010101 11101101. Una delle istruzioni primitive in ogni linguaggio macchina è un'istruzione di USCITA, che prende il contenuto attuale dell'accumulatore (per es., il numero binario 01100001) e lo scrive sullo schermo o la stampante, così che un utente esterno può accedere ai risultati raggiunti nella CPU. In una versione leggermente più amichevole nei confronti dell'utente, una procedura composta di una serie di istruzioni primitive può prima tradurre il numero binario nella notazione decimale (per es., binario 00000110 = decimale 6) o in una lettera dell'alfabeto tramite il codice ASCII (per es., binario 01100001 = «a» e 01000001 = «A») e poi stampare il risultato. Queste sottoprocedure sono alla base delle più elaborate istruzioni di uscita che troviamo nei linguaggi di programmazione di più alto livello, come il FORTRAN o il Pascal o il Lisp. Ciò permette al programmatore di creare ulteriori sottoprocedure per costruire messaggi più complessi, andando a prelevare lunghe serie di numeri nella memoria, farli passare nell'accumulatore, tradurli e scrivere il risultato sullo schermo o la stampante. Per esempio, una sottoprocedura può andare varie volte all'accumulatore per prendere i valori da inserire negli spazi vuoti di

Buon __________, Sig. ______________ il vostro scoperto è di lire __________. Arrivederci, Sig. ____________!

– uno schema «preconfezionato» di frase che viene essa stessa tenuta in memoria come una serie di numeri binari fino a che qualche sottoprocedura non stabilisce che è tempo di «aprire la confezione». In questa maniera, una precisa gerarchia di procedure stabilite può trasformare sequenze di contenuti specifici nell'accumulatore in espressioni che un essere umano può leggere sullo schermo o la stampante: «Vuoi salvare questo documento?» o «6 file copiati» o «Ciao, Giorgio, vuoi giocare a filetto?».

Ci sono due caratteristiche di questo processo che sono condivise dal modello di Levelt: (1) il processo prende come dato in ingresso un contenuto già determinato, e (2) la burocra-

zia – il «flusso di controllo» nel gergo informatico – deve essere stata progettata accuratamente: tutte le «autorizzazioni a decidere» provengono gerarchicamente dall'alto, grazie ad una delega della responsabilità ai subagenti le cui descrizioni dei compiti dettano quale porzione di analisi degli obiettivi e degli strumenti sono autorizzati ad eseguire. È piuttosto interessante che la prima di queste caratteristiche – il contenuto determinato – sembra essere suffragata dal parere di Otto sui suoi processi: c'è un determinato «pensiero» lì nel Centro, che aspetta di «essere messo in parole». La seconda caratteristica comune sembra però strana: la gerarchia di procedure che servilmente trasforma nel linguaggio naturale *quel particolare pensiero* è stata progettata anteriormente da *qualcun altro* – dal programmatore, nel caso della macchina di von Neumann, e presumibilmente da una combinazione dell'evoluzione e dello sviluppo individuale nel caso delle attività del Formulatore di Levelt. Il ruolo decisionale e creativo che il pensatore del pensiero dovrebbe svolgere nel mettere il pensiero in parole non appare nel modello; o è stato usurpato dal Concettualizzatore, che svolge tutto il lavoro creativo prima di inviare l'ordine al Formulatore, o è implicito nel progetto del Formulatore, un *fait accompli* di qualche precedente processo progettuale.

In quale altro modo potrebbero essere organizzati gli obiettivi e gli strumenti? Consideriamo una caricatura opposta: un pandemonio di folletti-parole. Ecco come parliamo: prima entriamo nella modalità «rumori vocali» – accendiamo l'altoparlante:

Beeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeeep....

Non c'è nessuna buona ragione per farlo, ma del resto in noi non si verifica nemmeno qualche buona ragione per non farlo. Il «rumore» interno eccita vari folletti in noi che cominciano a interferire con il suono dell'altoparlante cercando di modularlo in infiniti modi casuali. Ne risulta un farfugliare, ma almeno è un farfugliare italiano (per chi parla in italiano):

Para-paponzi-ponzi-pero-paralalà-lallà-nonnò...

Ma prima che la minima parte di questa roba imbarazzante si diffonda nel mondo esterno, altri folletti, sensibili alle configurazioni sonore nel caos, iniziano a forgiarle in parole, frasi, modi di dire...

Ah sì, e come mai? Il calcio, non sai, di fatti, fragole, coincidenze, va bene? Ecco il biglietto. E con questo?

il che incita i folletti a fare ulteriori sbalorditive scoperte, accresciute da un'agglomerazione opportunistica che conduce a una verbosità più accettabile, fino a che non salta fuori un'intera frase:

Ti faccio inghiottire i denti!

Fortunatamente, però, questa viene messa da parte, rimane non detta, poiché nello stesso tempo (in parallelo) altri candidati che erano in fabbricazione sono ora in vista, incluso qualche ovvio perdente, come

Sei un grande spilorcio!

e

Hai letto qualche buon libro ultimamente?

e un vincitore, in mancanza di concorrenti migliori, che viene pronunciato:

I tuoi piedi sono troppo grandi!

La musa non ha arriso al nostro parlante in questa occasione: nessuna frecciata sagace è arrivata in finale, ma almeno è stato propalato qualcosa di più o meno appropriato alla sua corrente «disposizione mentale». Mentre si allontana dopo l'incontro, probabilmente il parlante riprenderà il caotico torneo, borbottando e meditando su quello che avrebbe dovuto dire. La musa può allora discendere su di lui con qualcosa di meglio, ed egli lo assaporerà, rigirandolo varie volte nella sua mente, immaginando l'espressione ferita che avrebbe provocato sulla faccia del suo interlocutore. Prima di essere giunto a casa, il parlante potrà «ricordare» vividamente di aver infilzato il suo interlocutore con un'arguzia tagliente.

Possiamo supporre che tutto ciò avviene in rapide generazioni all'interno di un «dispendioso» processo in parallelo, con orde di anonimi folletti e di loro speranzose costruzioni che non vedono mai la luce del sole – né come opzioni che vengono *consciamente* scrutinate e rigettate, né come atti linguistici defini-

tivamente eseguiti per il beneficio di un pubblico esterno. Se si concede un tempo sufficiente, più di un'opzione può essere provata in una ripetizione cosciente, ma una simile audizione formale si verifica raramente, viene riservata alle occasioni in cui la posta in gioco è così alta da far temere che una parola sbagliata possa provocare pesanti penalità. Nei casi normali, il parlante non gode di nessuna anteprima; lui e il suo pubblico apprendono nello stesso tempo quali sono le parole pronunciate dal parlante.

Ma come viene giudicato questo torneo di parole? Quando una parola o una frase o un intero periodo batte il suo concorrente, come viene discriminata e valutata la sua adeguatezza o appropriatezza alla corrente disposizione mentale del parlante? Che cos'è una disposizione mentale (se non un'esplicita intenzione comunicativa) e come agisce nel torneo la sua influenza? Perché, dopo tutto, anche se non c'è un Autore Centrale deve esserci un modo in cui il contenuto passa dalle profondità interne del sistema – dai suoi processi percettivi, per esempio – ai resoconti verbali.

Ricapitoliamo il problema. Il guaio dell'estremo burocratico è che il Concettualizzatore sembra sinistramente potente, un homunculus con troppa conoscenza e responsabilità. Questo eccesso di potere è evidente nell'imbarazzante problema di come esprime il suo prodotto, il *messaggio preverbale*. Se specifica di già un atto linguistico – se è già una sorta di atto linguistico in mentalese, un determinato comando al Formulatore – la maggior parte del lavoro difficile avviene prima che il nostro modello compaia. Il guaio dell'alternativa pandemonio è che dobbiamo trovare un modo in cui le fonti del contenuto possano *influenzare* o *limitare* le energie creative dei folletti-parole senza dettar legge su di essi.

Che dire del processo descritto nel capitolo 1, quei giri di risposte a domande che generavano le allucinazioni sul modello del gioco della psicoanalisi? Ricorda che avevamo eliminato il saggio drammaturgo freudiano e il produttore di allucinazioni rimpiazzandoli con un processo dal quale il contenuto *emergeva* grazie ad un incessante interrogare di un interrogante. Il problema residuo era quello di trovare un modo per sbarazzarci dell'intelligente interrogante, un problema che avevamo posposto. Qui abbiamo il problema complementare: come ottenere le risposte ad un'impaziente folla di concorrenti che chiedono domande del genere «Perché non diciamo "Tua madre calza stivali militari!"?» o (in un altro contesto) «Perché non diciamo

"Mi sembra di vedere un punto rosso che si muove e diventa verde muovendosi"?». *Sono due problemi complementari – non si potrebbero forse risolvere vicendevolmente accoppiandoli?* E se i folletti-parole fossero, in parallelo, gli interroganti/concorrenti, e i folletti-contenuto fossero i rispondenti/giudici? Le intenzioni comunicative completamente sviluppate ed eseguite – i significati – potrebbero emergere da un processo quasi-evoluzionistico di progettazione di atti linguistici che implica la collaborazione, parzialmente seriale, parzialmente in parallelo, di vari sottosistemi, dei quali nessuno è in grado di eseguire – o ordinare – per proprio conto un atto linguistico.

Un processo del genere è realmente possibile? Ci sono svariati modelli di simili processi «ad adempimento di condizioni» ed hanno effettivamente dei poteri sorprendenti. Oltre alle varie architetture «connessioniste» di elementi simili ai neuroni (vedi, per es., McClelland e Rumelhart, 1986), ci sono altri modelli più astratti. L'architettura Jumbo di Douglas Hofstadter (1983), che cerca soluzioni per *Jumbles* o anagrammi, ha le caratteristiche giuste, come anche le idee di Marvin Minsky (1985) sugli Agenti che formano la «società della mente» – che sarà esaminata meglio nel capitolo 9. Ma dobbiamo astenerci dal giudicare fino a che non vengano creati e messi in funzione dei modelli che siano più dettagliati, espliciti e specificamente diretti alla produzione linguistica. Ci potrebbero essere sia sorprese che delusioni.

Noi sappiamo, comunque, che in qualsiasi modello riuscito della produzione linguistica dovremo in qualche modo utilizzare un processo evoluzionistico della generazione del messaggio, altrimenti saremo bloccati da un miracolo («E poi avviene un miracolo») o un regresso all'infinito di Autori per stabilire il compito.[1] Noi sappiamo anche – dalle ricerche che Levelt passa in rassegna – che ci sono dei processi piuttosto rigidi e automatici che entrano in gioco alla fine e determinano le trasformazioni, dal grammaticale al fonologico, che compongono le finali ricette muscolari per il parlare. Le due caricature definiscono gli estremi di un *continuum*, dall'iperburocratico all'ipercaotico. Il vero modello di Levelt – non la caricatura che ne ho dato per far risaltare meglio il contrasto – include (o facilmente si può fare in modo che includa) alcune delle caratteristiche non burocratiche della caricatura opposta: per esempio, non c'è nulla di profondo o strutturale che impedisca al Formulatore di Levelt di imbarcarsi in una generazione linguistica più o meno spontanea (non richiesta, non diretta) e, dato il ciclo di con-

trollo che attraverso il Sistema di Comprensione del Discorso ritorna al Concettualizzatore (vedi fig. 8.1), questa attività spontanea *potrebbe* giocare proprio quella sorta di ruolo generativo immaginato per i molteplici folletti-parole. Tra le due caricature c'è uno spettro di modi più realistici in cui i modelli alternativi potrebbero essere sviluppati. La questione principale è quanta interazione ci sia tra gli specialisti che determinano contenuto e stile di ciò che deve essere detto e gli specialisti che «conoscono le parole e la grammatica».

Ad un estremo, la risposta è: nessuna. Potremmo mantenere il modello di Levelt intatto e integrarlo semplicemente con un modello in stile pandemonio di ciò che avviene *all'interno* del Concettualizzatore per risolvere il problema della generazione del «messaggio preverbale». Nel modello di Levelt, c'è una separazione quasi completa tra i processi che generano il messaggio (definizione dell'oggetto) e la produzione linguistica (realizzazione dell'oggetto). Quando la prima porzione del messaggio preverbale arriva al Formulatore, scatta la produzione della prima parte dell'espressione linguistica; il modo in cui le parole vengono scelte dal Formulatore limita le possibili continuazioni dell'espressione, ma c'è solo una minima *collaborazione* sulla revisione della definizione dell'oggetto. Gli esecutori materiali del linguaggio, i dipendenti che compongono il Formulatore, sono, nei termini di Jerry Fodor, «incapsulati»; nella loro maniera automatica, fanno del loro meglio con gli ordini che ricevono, senza «se», «e» o «ma».

All'estremo opposto ci sono i modelli in cui le parole e le frasi, partendo dal Lessico, assieme ai loro suoni, significati e associazioni, sgomitano in un pandemonio con le costruzioni grammaticali per aprirsi una strada di accesso al messaggio di cui tutte vogliono far parte, e alcune di loro danno in questo modo un contributo sostanziale alle intenzioni comunicative stesse, che però solo poche di loro finiscono per mettere in atto. A questo estremo, le intenzioni comunicative esistenti sono tanto una causa quanto un effetto del processo – emergono come un prodotto, e, una volta emerse, possono fungere da paradigmi su cui misurare *ulteriori* implementazioni delle intenzioni. Non c'è un'unica fonte del significato, ma molte fonti che si alternano, sviluppate opportunisticamente dalla ricerca per le parole giuste. Invece di un determinato contenuto situato in un particolare luogo funzionale e che attende di essere «italianizzato» da specifiche sottoprocedure, c'è una disposizione mentale non ancora completamente determinata e distribuita in

vari luoghi del cervello che incanala il processo compositivo, il quale a sua volta può nel corso del tempo retroagire per effettuare aggiustamenti e correzioni e specificare ulteriormente i compiti espressivi che misero in moto il processo compositivo in origine. C'è ancora una configurazione generale di passaggio seriale, con concentrazione su un argomento alla volta, ma i confini non sono linee nette.

Nel modello stile pandemonio, il controllo viene usurpato piuttosto che delegato, in un processo che è ampiamente non finalizzato e opportunistico; esistono fonti multiple per le «decisioni» progettuali che conducono all'espressione linguistica finale, e non è possibile una divisione netta tra gli ordini relativi al contenuto provenienti dall'interno e i suggerimenti per la realizzazione offerti volontariamente dai folletti-parole. I modelli di questo tipo suggeriscono che per *conservare* il ruolo creativo di «colui che *esprime* il pensiero» (qualcosa che stava molto a cuore a Otto), dobbiamo *abbandonare* l'idea che «colui che *pensa* il pensiero» parta da un determinato pensiero che deve essere espresso. Anche questa idea di un contenuto determinato stava molto a cuore ad Otto, ma qualcosa deve pur concedere (il paragrafo 3 indagherà le alternative in maggior dettaglio).

In quale punto dello spettro risiede la verità? Questa è una domanda empirica di cui ancora non conosciamo la risposta.[2] Ci sono alcuni fenomeni, comunque, che spingono fortemente a pensare (me, perlomeno) che la generazione linguistica debba implicare una presenza diffusa di processi – opportunistici, paralleli, evoluzionistici – in stile pandemonio. Nel prossimo paragrafo passerò rapidamente in rassegna alcuni di essi.

3. *Quando le parole vogliono essere dette*

> Qualunque cosa vogliamo dire, è probabile che ciò che diremo non sarà esattamente *quello*.
>
> MARVIN MINSKY (1985), p. 460

I ricercatori dell'IA Lawrence Birnbaum e Gregg Collins (1984) hanno evidenziato un aspetto singolare dei lapsus freudiani. Come è noto Freud attrasse la nostra attenzione sui *lapsus linguae* che non sono – secondo lui – casuali o senza significato, ma profondamente significativi: inserzioni inconsciamente *volute* nel tessuto del discorso, inserzioni che soddisfano indirettamente o parzialmente i desideri comunicativi repressi del par-

lante. Questa tipica tesi freudiana è stata spesso rifiutata con veemenza dagli scettici, ma c'è un problema interessante nella sua applicazione a casi particolari che non hanno nulla a che fare con l'opinione che si può avere sui temi freudiani più oscuri come la sessualità, il complesso di Edipo o gli impulsi di morte. Freud discute un esempio in cui un uomo disse:

Signori, vi invito ad *avere il singhiozzo* alla salute del nostro Capo.

(In tedesco – la lingua effettivamente parlata nell'esempio – la parola per «avere il singhiozzo», *aufzustossen*, fu inserita al posto della parola per «brindare», *anzustossen*).

Nella sua spiegazione, Freud sostiene che questo lapsus è una manifestazione del desiderio inconscio del parlante di ridicolizzare o insultare il suo superiore, soppresso dall'obbligo sociale e politico di onorarlo. Comunque, [...] non possiamo ragionevolmente aspettarci che l'intenzione del parlante di ridicolizzare il suo superiore abbia fatto sorgere fin dall'inizio un piano che implicasse l'uso della parola «avere il singhiozzo»: a *priori*, ci sono centinaia di parole e frasi che possono essere usate più plausibilmente per insultare o ridicolizzare qualcuno [..]. In nessuna maniera un pianificatore avrebbe potuto ragionevolmente anticipare che il desiderio di ridicolizzare o insultare il suo superiore sarebbe stato soddisfatto dicendo la parola «avere il singhiozzo», proprio perché in primo luogo sembra del tutto implausibile che il pianificatore avrebbe scelto di usare quella parola come un insulto.

L'unico processo che potrebbe spiegare la frequenza di incredibili lapsus freudiani, essi sostengono, è uno di «pianificazione opportunistica».

[...] Ciò che esempi come quello precedente sembrano indicare, quindi, è che i desideri *stessi* siano attivi agenti cognitivi, capaci di dirigere le risorse cognitive necessarie per riconoscere le opportunità che possono soddisfarli e le risorse comportamentali necessarie per sfruttare le opportunità. (Birnbaum e Collins, 1984, p. 125)

I lapsus freudiani sono interessanti perché sembrano nello stesso tempo errori e non-errori, ma il fatto (se è così) che soddisfano desideri *inconsci* non rende più difficile la loro spiegazione di quella di altre scelte di parole che esaudiscono contemporaneamente varie funzioni (o desideri). È altrettanto difficile

immaginare come i vari tipi di giochi di parole possano essere il risultato di una pianificazione e produzione non-opportunistiche, incapsulate. Se qualcuno conosce un metodo per produrre arguzie – un metodo dettagliato che realmente funziona – si faccia avanti, perché ci sono molti commediografi che lo pagherebbero fior di quattrini.[3]

Se Birnbaum e Collins hanno ragione, l'uso creativo del linguaggio può essere realizzato solo da un processo parallelo in cui molteplici *obiettivi* o *desideri* siano simultaneamente in allerta per il materiale. Ma se i materiali stessi fossero contemporaneamente in allerta per le occasioni di essere incorporati? Noi prendiamo il nostro vocabolario dalla nostra cultura; le parole e le frasi costituiscono le caratteristiche fenotipiche più salienti – i corpi visibili – dei memi che ci invadono, ed è difficile che ci sia un mezzo in cui i memi possano replicarsi più congeniale di un sistema di produzione linguistica in cui i burocrati supervisori abbiano parzialmente abdicato, cedendo gran parte del controllo alle parole stesse, che difatti combattono tra di loro per ottenere un'opportunità di comparire nella ribalta dell'espressione pubblica.

Non è una novità che talvolta diciamo qualcosa soprattutto perché ci piace il modo in cui suona, non perché ci piace il suo significato. Nuove espressioni gergali si diffondono attraverso i gruppi sociali, insinuandosi nel linguaggio di quasi tutti, anche di quelli che tentano di resistere. Pochi tra quelli che usano una nuova parola stanno seguendo deliberatamente o consciamente la massima del maestro di scuola: «Usa una parola nuova tre volte e sarà tua!». Ad un livello più ampio di aggregazione, intere frasi ci piacciono per il modo in cui risuonano nel nostro orecchio o saltellano sulla nostra lingua, indipendentemente dalla loro più o meno effettiva corrispondenza alle descrizioni proposizionali che avevamo già deciso. Una delle espressioni più citate di Abraham Lincoln è:

Puoi ingannare tutti qualche volta e qualcuno tutte le volte, ma non puoi ingannare tutti tutte le volte.[4]

Che cosa intendeva dire Lincoln? Gli studiosi di logica si sono divertiti a sottolineare che c'è una «ambiguità di sfera semantica» in questa frase. Lincoln intendeva asserire che c'è qualche somaro che può sempre essere ingannato oppure

che in ogni occasione c'è qualcuno o qualcun altro destinato ad essere ingannato – ma non sempre la stessa persona? Logicamente, queste sono proposizioni completamente differenti.

Fai il paragone con:

«Qualcuno vince sempre la lotteria»
«Deve essere truccata!»
«Non intendevo questo»

Quale interpretazione Lincoln aveva in testa? Forse nessuna delle due! Che probabilità ci sono che Lincoln non si sia mai accorto dell'ambiguità semantica e non abbia mai effettivamente avuto un'intenzione comunicativa piuttosto che «l'altra»? Forse la frase gli è semplicemente suonata così bene la prima volta che l'ha formulata che non ha mai colto l'ambiguità, *e non ha mai avuto delle precedenti intenzioni comunicative* – tranne l'intenzione di dire genericamente qualcosa di conciso e ben cadenzato sulla possibilità di ingannare la gente. Le persone parlano in questa maniera, anche i grandi Autori come Lincoln.

La narratrice Patricia Hampl, nel suo penetrante saggio *The Lax Habits of the Free Imagination*, scrive sui propri processi di composizione di racconti:

Ogni storia ha una storia. Questa storia segreta, che ha una scarsa probabilità di essere raccontata, è la storia della sua creazione. Forse la «storia della storia» *non può* mai essere raccontata, giacché un lavoro finito distrugge la propria storia, la rende obsoleta, una scoria. (Hampl, 1989, p. 37)

Il lavoro finito, lei nota, è facilmente interpretabile dai critici come un artefatto astutamente escogitato per soddisfare una schiera di sofisticate intenzioni dell'autore. Ma quando incontra queste ipotesi sul suo proprio lavoro, lei diventa imbarazzata:

«Hampl» aveva ben poche intenzioni, eccetto quella, come il ciarlatano in cui immediatamente mi sono immedesimata, di sgraffignare qualunque cosa disponibile sul mio tavolo che si adattasse ai miei scopi immediati. Il peggio è che tali «scopi» erano vaghi, incoerenti, mutevoli, sotto pressione. E chi – o che cosa – stava facendo pressione? Non potrei dirlo. (p. 37)

Come procede lei, allora? Hampl suggerisce una massima: «Continua a parlare – borbottare è utile». Alla fine, i borbottii prendono una forma che incontra l'approvazione dell'autore. Il processo che Hampl individua in grande scala nella sua scrittura creativa non potrebbe essere un semplice ingrandimento di un processo più sommerso e più veloce che produce il parlare creativo nella vita di tutti i giorni?

L'allettante somiglianza non implica solo un processo ma anche una reazione o un'attitudine conseguente. Lo zelo confessionale di Hampl contrasta con una reazione più normale – e non totalmente disonesta – che gli autori hanno di fronte alle interpretazioni amichevoli da parte dei lettori: questi autori si rimettono garbatamente all'attribuzione di intenti e sono anche disposti ad approfondirli, dicendo cose simili a «Già, suppongo che questo sia quello che intendevo, fin dall'inizio!». E perché no? È forse contraddittorio pensare che una mossa che si è appena fatta (negli scacchi, nella vita, nella scrittura) sia realmente più astuta di quella che si era dapprincipio creduto? (Per ulteriori riflessioni su questo argomento, vedi Eco, 1990.)

Come E.M. Forster disse «Come posso dire che cosa penso finché non vedo che cosa dico?» noi scopriamo cosa pensiamo (e quindi che cosa intendiamo) riflettendo su quello che diciamo – e non correggiamo. Così noi siamo, almeno in quelle occasioni, sulla stessa barca dei nostri critici ed interpreti esterni, incontriamo un testo e cerchiamo di ricavarne la migliore lezione possibile. Il fatto che noi lo abbiamo detto gli dà una certa persuasività personale o almeno una presupposizione di autenticità. Se lo ho detto (e ho udito me stesso dirlo e non ho udito me stesso affrettarsi a correggerlo), *probabilmente*, intendevo dirlo, e probabilmente significa ciò che sembra significare – a me.

La vita di Bertrand Russell ci offre un esempio:

Si fece molto tardi prima che i due ospiti se ne andassero e Russell rimase solo con Lady Ottoline. Rimasero seduti a parlare davanti al fuoco fino alle quattro del mattino. Russell, ripensando all'evento alcuni giorni più tardi, scrisse: «Non sapevo di amarti fino a che non ho udito me stesso dirtelo – per un istante pensai "Buon Dio, cosa ho detto?" e poi seppi che era la verità». (Clark, 1975, p. 176)

Che dire, però, delle altre occasioni in cui non abbiamo un tale senso di *scoperta*, di auto-interpretazione rivelatrice? Potremmo supporre che in questi casi, normali, noi godiamo in

anticipo di un qualche sguardo intimo e privilegiato su ciò che intendiamo, semplicemente perché noi siamo gli Autori, la *fons et origo* del significato delle parole che *noi* diciamo, ma una tale supposizione si deve basare su un argomento e non sull'appello alla tradizione. Potrebbe essere benissimo che non abbiamo alcun senso di scoperta in questi casi solo perché ci appare talmente ovvio ciò che noi intendiamo. Non serve nessun «accesso privilegiato» per intuire che quando alla tavola da pranzo io dico «Per favore, passami il sale», io intendo chiedere il sale.

Un tempo credevo che non ci fossero alternative ad un Autore Centrale, ma pensavo di avergli trovato un rifugio sicuro. In *Contenuto e coscienza*, sostenevo che ci deve essere una linea funzionalmente rilevante (che chiamavo linea della consapevolezza) che separa la definizione preconscia delle intenzioni comunicative dalla loro successiva esecuzione. Potrebbe essere incredibilmente difficile localizzare, anatomicamente, questa linea nel cervello, perché potrebbe essere estremamente contorta, ma essa deve esistere, logicamente, come lo spartiacque che divide le disfunzioni in due varietà. Gli errori potrebbero verificarsi ovunque nell'intero sistema, ma ogni errore deve cadere – per necessità geometrica – da una parte o dall'altra della linea. Se cadessero nella parte dell'esecuzione, sarebbero (correggibili) errori di *espressione*, come i *lapsus linguae*, le papere e altri errori di pronuncia. Se cadessero nella parte più interna o più alta, essi *modificherebbero quello che sta per essere espresso* (il «messaggio preverbale» nel modello di Levelt). Il significato viene stabilito su questo spartiacque; è lì che il significato nasce. Ci deve essere un luogo del genere dove il significato nasce, io pensavo, poiché *qualcosa* deve fissare lo standard rispetto al quale la «retroazione» può rilevare un fallimento nell'esecuzione.

Il mio errore era dovuto alla stessa ambiguità di sfera semantica che rendeva incerta l'interpretazione della massima di Abe Lincoln. Deve esserci effettivamente qualcosa che in ogni occorrenza costituisce, per l'occasione, lo standard in base al quale ogni «errore» corretto viene corretto, ma non ci deve essere la stessa singola cosa ogni volta – anche *all'interno* di un singolo atto linguistico. Non deve esserci una linea *fissa* (sia pur di incerta localizzazione) che segna questa distinzione. In realtà, come abbiamo visto nel capitolo 5, la distinzione tra revisioni pre-esperienziali che *cambiano ciò che costituiva l'esperienza* e le revisioni post-esperienziali che hanno l'effetto di

modificare il resoconto o il ricordo di ciò che costituiva l'esperienza è indeterminata in situazioni limite. Talvolta i soggetti sono spinti a rivedere o emendare le loro asserzioni, e talvolta non lo sono. Talvolta quando apportano effettivamente alcune revisioni, le versioni riviste non sono più vicine alla «verità» o a «ciò che intendevano *realmente* dire» di quelle rimpiazzate. Come abbiamo già notato, il momento in cui la revisione editoriale precedente alla pubblicazione lascia il posto all'inserzione degli errata posteriore alla pubblicazione è definibile solo in modo arbitrario. Quando domandiamo ad un soggetto se una particolare confessione pubblica cattura o no adeguatamente l'ultima verità interiore su ciò che costituiva la sua esperienza, egli non si trova in una posizione per giudicare migliore della nostra di estranei. (Vedi anche Dennett, 1990d.)

Ecco un altro modo di vedere lo stesso fenomeno. Ogni qualvolta si verifica un processo di creazione di un'espressione verbale, c'è all'inizio una distanza che deve essere eliminata: la «distanza del disaccordo nello spazio semantico», potremmo dire, tra il contenuto che deve essere espresso e i vari candidati per l'espressione che sono inizialmente proposti. (Secondo la mia vecchia posizione, questo era un problema di semplice «correzione retroattiva» con un *punto stabilito* che fungeva da standard sul quale i candidati verbali venivano misurati, scartati, migliorati.) Il processo che fa la spola tra i due estremi della distanza, e che così facendo la riduce, è una sorta di processo retroattivo, ma è possibile sia che l'espressione proposta venga sostituita o revisionata per meglio adattarsi al «contenuto che deve essere espresso», sia che il «contenuto che deve essere espresso» si modifichi per adattarsi meglio a qualche espressione proposta. In questo modo, le parole o le frasi più accessibili o disponibili potrebbero effettivamente *modificare il contenuto dell'esperienza* (se intendiamo l'esperienza come ciò che alla fine viene riferito – un contenuto assodato nel mondo eterofenomenologico del soggetto).[5]

Se la nostra unità come Autori non è meglio garantita di così, allora dovrebbe essere possibile, in linea di principio, che essa venga frantumata in qualche rara occasione. Ecco due casi in cui ciò sembra essere avvenuto.

Un giorno mi fu chiesto insistentemente di fare l'arbitro di prima base in una partita di baseball – un compito nuovo per me. Nell'ultimo momento cruciale della partita mi è toccato decidere lo status del battitore che correva in prima base. Era una decisione difficile e mi sono accorto che gridavo «SALVO!»

mentre sollevavo enfaticamente il pollice verso l'alto – il segnale per FUORI. Nel tumulto susseguente mi è stato chiesto di chiarire che cosa intendevo. Onestamente non potevo dirlo, almeno non da una posizione privilegiata. Infine decisi (con me stesso) che poiché non avevo molta pratica con i segnali manuali, ma ero un competente utilizzatore di parole, il mio atto vocale doveva avere la meglio, ma chiunque altro avrebbe potuto fare lo stesso ragionamento. (Sarei felice di conoscere altri aneddoti in cui le persone si sono trovate a non sapere quale di due differenti atti linguistici intendevano eseguire.)

In una situazione sperimentale, lo psicologo Tony Marcel (in stampa) si è imbattuto in un caso ancora più drammatico. A un soggetto, affetto da visione cieca (su cui mi dilungherò nel capitolo 11), fu chiesto di *dire* ogni qualvolta pensava che ci fosse un lampo luminoso, ma gli furono date precise istruzioni su come doveva farlo. Gli fu chiesto di eseguire questo singolo atto linguistico tramite tre differenti atti contemporanei (non in sequenza, ma neppure necessariamente «all'unisono»):

1. dire «sì»,
2. premere un pulsante (il pulsante del sì)
3. accennare sì con il capo.

La cosa sorprendente è che il soggetto non eseguiva sempre tutti e tre gli atti insieme. Talvolta capitava che accennasse di sì, ma non diceva sì o non premeva il pulsante del sì, o simile. Non c'era un modo diretto per catalogare le tre risposte differenti, né in base alla fedeltà all'intenzione né in base all'accordo con l'evento. Cioè, quando sussisteva un disaccordo tra le tre azioni, il soggetto non aveva un modo organico per stabilire quale atto dovesse essere accettato e quale dovesse essere considerato un lapsus della lingua, del dito o del capo.

Rimane da vedere se è possibile provocare casi simili in altre situazioni con altri soggetti, normali o no, ma anche altre condizioni patologiche suggeriscono un modello di produzione linguistica in cui la verbalizzazione *può* essere messa in moto senza nessun ordine che proviene dall'Autore Centrale. Se soffri di una di queste patologie, «la tua mente è in vacanza, ma la tua bocca fa gli straordinari», come dice la canzone di Mose Allison.

L'*afasia* è una perdita o un danno della capacità di parlare;

molte varietà di afasia sono piuttosto comuni e sono state ampiamente studiate da neurologi e linguisti. Nel tipo più comune, l'afasia di Broca, il paziente è profondamente consapevole del problema e lotta, con crescente frustrazione, per cercare le parole che gli sfuggono dalla punta della lingua. Nell'afasia di Broca, l'esistenza di intenzioni comunicative frustrate è dolorosamente chiara al paziente. Ma in un tipo relativamente raro di afasia, l'afasia gergale o fluente, i pazienti non sembrano avere la minima ansietà sulle loro deficienze verbali.[6] Anche se sono di intelligenza normale, assolutamente non psicotici o dementi, sembrano del tutto soddisfatti di esecuzioni verbali simili a queste (tratte da due casi descritti da Kinsbourne e Warrington, 1963):

1° caso:

Come sta oggi?
«Pettegolmente bene e i Lords e il cricket e le battaglie di Scozia e di Inghilterra. Non saprei. Ipertensione e due cricket vinti, servire, battere e afferrare, poveretti, annullamenti forse pettegolezzi, annullamenti, braccio ed argomento, finire di servire».
Qual è il significato di «sicurezza innanzitutto»?
«Guardare e vedere e in particolare la Richmond Road, e guardare il traffico ed esitazione a destra e passeggiare, una causa molto buona, forse, le zebre forse queste, autovetture e i semafori».

2° caso:

Lavorava in un ufficio?
«Sì, lavoravo in un ufficio».
E che tipo di ditta era?
«Oh, come un funzionario di questa, e la lamentela era discutere le tonazioni di ciò che era stampato, come era stampato, e impedire che differenti ... rtico ... rticoli, prendermi dall'attributo convertsamenti ... scusi ...»

«Lei vuole dare la vocazione soggettiva per mantenere la vocazione di impregnazione consanguana perfetta.»

«La sua normale corrucazione sarebbe un punto.»

Alla richiesta di identificare una lametta per unghie:

«È un coltello, una coda di coltello, un coltello, caduto, un coltello caduto.»

e le forbici:

«Cannali – è un cannali – non veramente un cannali – due cannali che contengono un pettine – no, non un pettine – due cannali che servono ad impedire che il comandante non sia ora».

Una condizione stranamente simile, e molto più comune, è la *confabulazione*. Nel capitolo 4, ho suggerito che la gente normale può spesso confabulare riguardo ai dettagli della propria esperienza, poiché tende ad indovinare senza accorgersene, e scambia la teorizzazione per osservazione. La confabulazione patologica implica delle invenzioni inconsapevoli di un ordine completamente differente. Spesso nei casi di danno cerebrale, specialmente quelli che comportano una tremenda perdita di memoria – come nella sindrome di Korsakoff (una tipica conseguenza dell'alcolismo grave) – le persone continuano a cianciare affermando totali falsità sulla loro vita e le loro storie passate, e perfino su eventi degli ultimi minuti, se la loro amnesia è grave.

Il cicaleccio che ne risulta sembra virtualmente normale. Assomiglia infatti alle poco interessanti chiacchiere stereotipate che passano per una conversazione da bar: «Oh, sì, mia moglie ed io – siamo stati nella stessa casa per trent'anni – avevamo l'abitudine di andare a Coney Island e sederci sulla spiaggia – ci *piaceva moltissimo* starcene lì seduti, a guardare i giovani e ..., ma questo era prima dell'incidente». Ma è tutto inventato di sana pianta. Sua moglie può essere morta molti anni fa, forse non hanno mai abitato entro un raggio di cento chilometri da Coney Island e si sono spostati di appartamento in appartamento. Spesso un ascoltatore sprovveduto potrebbe benissimo non accorgersi di trovarsi di fronte a un confabulatore, infatti i ricordi sembrano così naturali e «sinceri» e le risposte così pronte.

I confabulatori non sono consapevoli del fatto che stanno inventando tutto e gli afasici gergali ignorano di confezionare una semplice macedonia di parole. Queste sbalorditive anomalie sono esempi di *anosognosia* o inabilità a riconoscere un deficit. Esistono anche altri tipi di una simile incapacità a comprendere le proprie carenze e nel capitolo 11 vedremo cosa possono dirci sull'architettura funzionale della coscienza. Nel frattempo, possiamo notare che il meccanismo cerebrale è in grado di costruire

degli apparenti atti linguistici senza che ci sia una coerente direzione proveniente dall'alto.[7]

La patologia, sia la temporanea deformazione indotta da ingegnosi esperimenti, sia le più gravi e permanenti alterazioni causate da malattie o danni meccanici al cervello, fornisce molti suggerimenti su come è organizzato il macchinario. Questi fenomeni mi inducono a pensare che la nostra seconda caricatura, il Pandemonio, sia più vicina alla verità di quanto non lo sia il modello più rispettabile e burocratico, ma ciò deve essere sottoposto a un appropriato esame empirico. Non voglio sostenere che sia impossibile per un modello sostanzialmente burocratico rendere giustizia a queste patologie, ma semplicemente che esse non sembrerebbero i fallimenti naturali di un simile sistema. Nell'Appendice B, per gli scienziati, indicherò alcune direzioni di indagine che potrebbero aiutare a confermare o refutare le mie impressioni.

In questo capitolo io ho abbozzato – non certo dimostrato – una maniera in cui un fiume tumultuoso di prodotti verbali che sgorga da temporanee coalizioni formate da migliaia di folletti produttori di parole possa mostrare un'unità, l'unità di un'interpretazione che si ottimizza evolvendosi, che fa apparire tali folletti-parole *come se* fossero la realizzazione delle intenzioni di un Concettualizzatore – ed infatti lo sono, ma non di un Concettualizzatore *interno* che sia una parte propria di un sistema di produzione linguistica, ma del Concettualizzatore globale, la persona, di cui il sistema di produzione linguistica è esso stesso una parte propria.

Questa idea può dapprincipio sembrare astrusa, ma non dovrebbe sorprenderci. In biologia, abbiamo imparato a resistere alla tentazione di spiegare la *progettualità negli organismi* postulando una singola grande Intelligenza che fa tutto il lavoro. In psicologia, abbiamo imparato a resistere alla tentazione di spiegare il *vedere* dicendo che è come se ci fosse un osservatore interno seduto davanti a uno schermo, giacché l'osservatore interno farebbe tutto il lavoro – e se così fosse, il collegamento tra un tale homunculus e gli occhi sarebbe solo una sorta di cavo Tv. Analogamente dobbiamo sforzarci di resistere alla tentazione di spiegare l'*azione* come un qualcosa che sorge dagli imperativi di un comandante interno che svolge una parte troppo ampia del lavoro di pianificazione. Come al solito, il modo per sbarazzarci di un'in-

telligenza troppo grande per la nostra teoria è di sostituirla con una struttura, in ultima analisi meccanica, di semi-intelligenze semi-indipendenti che agiscono di concerto.

Questa considerazione non si applica solo alla generazione di atti linguistici, ma a qualsiasi azione intenzionale in generale. (Vedi Pears, 1984, per uno sviluppo di idee simili.) E, contrariamente alle prime apparenze, la fenomenologia ci aiuta in realtà a capire che le cose stanno così. Benché siamo talvolta coscienti di eseguire elaborati ragionamenti pratici, che conducono ad una conclusione su ciò che, tutto considerato, dovremmo fare, e che sono seguiti da una decisione cosciente di fare proprio quella cosa e dalla sua effettiva realizzazione finale, queste sono esperienze relativamente rare. La maggior parte delle nostre azioni intenzionali sono eseguite senza nessun preambolo del genere; e questo è un bene, poiché ce ne mancherebbe il tempo. L'errore comune è di supporre che questi casi relativamente rari di ragionamento pratico cosciente costituiscano un buon modello per il resto, i casi in cui le nostre azioni intenzionali emergono da processi ai quali non abbiamo accesso. Le nostre azioni di solito ci soddisfano; noi riconosciamo che sono sostanzialmente coerenti e che, per quanto ne sappiamo, contribuiscono in modo appropriato e tempestivo ai nostri progetti. Così non ci sembra errato presupporre che siano il prodotto di processi affidabili nel connettere strumenti ed obiettivi; cioè, sono razionali, in *un* senso della parola (Dennett, 1987a, 1991a). Ma questo non significa che sono razionali in un senso più stretto: il prodotto di un ragionamento seriale. Non dobbiamo spiegare i sottostanti processi sul modello di un ente interno che ragiona, conclude e decide accoppiando metodicamente strumenti e obiettivi e poi ordina l'azione particolare; abbiamo visto schematicamente un tipo differente di processo che potrebbe controllare il parlare, ed anche le altre nostre azioni intenzionali.

Lentamente ma bene, stiamo abbandonando le nostre cattive abitudini di pensiero e le stiamo sostituendo con altre abitudini. La scomparsa dell'Autore Centrale significa più in generale la scomparsa del Intendente Centrale (colui che ha intenzioni), ma il Boss vive ancora sotto altre spoglie. Nel capitolo 10 lo incontreremo nei ruoli dell'Osservatore e del Reporter (colui che riferisce), e dovremo trovare altri modi di pensare a ciò che accade, ma dobbiamo prima consolidare le fondamenta delle nostre nuove abitudini di pensiero legandole più strettamente a qualche dettaglio scientifico.

9

L'ARCHITETTURA DELLA MENTE UMANA

1. *Dove siamo?*

La parte difficile è finita, ma c'è ancora molto lavoro da fare. Il nostro duro allenamento per allargare l'orizzonte della nostra immaginazione può essere considerato concluso e ora siamo pronti a mettere alla prova la nostra nuova prospettiva. Lungo la strada siamo stati costretti a lasciare in sospeso molti problemi e a tollerare molte imprecisioni. Ci sono promesse da mantenere, debiti da riconoscere e paragoni da sviluppare. La teoria che sto proponendo include elementi tratti da molti pensatori. Io ho talvolta ignorato deliberatamente ciò che questi pensatori considerano le parti migliori delle loro teorie e ho messo insieme idee tratte da campi «ostili», ma la soppressione di questo ammasso di dettagli è stata dettata da un desiderio di chiarezza e d'incisività. Molti seri costruttori di modelli della mente possono essersi spazientiti per questo, ma non mi sembra che ci fosse nessun'altra maniera per guidare contemporaneamente vari tipi differenti di lettori al nuovo punto di vista. Ora, però, possiamo fermarci per fare l'inventario e consolidare alcuni dettagli essenziali. Il motivo che ci ha spinto ad affrontare tutte le difficoltà nella costruzione di una nuova prospettiva è, dopo tutto, quello di vedere in una nuova maniera i fenomeni e le controversie collegate. Così, diamo uno sguardo alla situazione.

In uno «schizzo in miniatura», ecco la mia teoria fino ad ora:

Non c'è un unico e definitivo «flusso di coscienza», perché non c'è un Quartier Generale centrale, un Teatro Cartesiano dove «tutto converge» per essere attentamente scrutinato da un Autore Centrale. Invece di un unico (per quanto ampio) flusso del genere, ci sono canali multipli in cui vari circuiti specializzati tentano, in un pande-

monio parallelo, di fare varie cose, creando man mano delle Molteplici Versioni. La maggior parte di queste frammentarie versioni di «narrazioni» giocano dei ruoli effimeri nella modulazione dell'attività in corso, ma qualcuna viene promossa ad ulteriori ruoli funzionali, in rapida successione, dall'attività di una macchina virtuale nel cervello. La serialità di questa macchina (il suo carattere «neumanniano») non è una caratteristica progettuale «cablata rigidamente», ma piuttosto il risultato di una successione di coalizioni di questi specialisti.

Gli specialisti fondamentali sono parte della nostra eredità animale. Non furono sviluppati per eseguire azioni specificamente umane, come leggere e scrivere, ma per evitare i predatori, schivare gli oggetti, riconoscere i volti, afferrare, scagliare, raccogliere le bacche e altri compiti essenziali. Spesso sono opportunisticamente trasferiti in nuovi ruoli, ai quali sono più o meno portati a seconda dei loro talenti originali. Non ne risulta un manicomio solo perché le tendenze imposte a tutta questa attività sono esse stesse il prodotto di un progetto. Parte di questo progetto è innato ed è condiviso con altri animali. Ma viene integrato, e spesso sopraffatto, da micro-abitudini di pensiero che vengono sviluppate nell'individuo, in parte come risultati di auto-esplorazione propri dell'individuo stesso e in parte come doni predefiniti della cultura. Migliaia di memi, nella maggior parte trasportati dal linguaggio, ma anche da «immagini» senza parole e altre strutture di dati, stabiliscono la loro residenza in un cervello individuale, forgiando le sue tendenze e trasformandolo quindi in una mente.

Questa teoria contiene abbastanza novità da rendere difficile una sua comprensione immediata, ma probabilmente perché attinge a modelli sviluppati da ricercatori in vari campi: psicologia, neurobiologia, Intelligenza Artificiale, antropologia – e filosofia. Questo eclettismo non viene scosso dal sospetto con cui viene visto frequentemente dai vari addetti ai lavori. Essendo un intruso abituale in questi campi, mi sono abituato alla sufficienza con cui alcuni ricercatori trattano i loro colleghi di altre discipline. «Perché, Dan», chiedono quelli dell'Intelligenza Artificiale, «sprechi il tuo tempo a parlare con i neuroscienziati? Non danno importanza ai "processi informazionali" e si preoccupano solo di *dove* essi avvengano, e di quali neurotrasmettitori siano implicati e di tutti questi fatti noiosi, ma non hanno la minima idea sui requisiti computazionali delle funzioni cognitive superiori». «Perché», domandano i neuroscienziati, «sprechi il tuo tempo con le fantasie dell'Intelligenza Artificiale? Quelli non fanno altro che inventare innumerevoli marchingegni e affermano cose di un'ignoranza imperdonabile sul cervello». Gli psicologi cognitivi, nel frattempo, sono accu-

sati di mettere assieme dei modelli privi *sia* di plausibilità biologica *sia* di dimostrati poteri computazionali; gli antropologi non riconoscerebbero un modello anche se lo vedessero, e i filosofi, come tutti sanno, non fanno altro che riciclare i panni sporchi degli altri, mettendo in guardia da confusioni che essi stessi hanno creato, in un'arena priva sia di dati che di teorie empiricamente verificabili. Con tutti questi idioti che lavorano al problema, non stupisce che la coscienza sia ancora un mistero.

Tutte queste accuse sono vere, e ce ne sarebbero altre, ma io non ho ancora incontrato alcun idiota. La maggior parte dei teorici dai quali ho attinto mi hanno colpito per la loro intelligenza – spesso sono brillanti, e con l'arroganza e l'impazienza che spesso accompagnano un'intelligenza brillante – ma con dei piani di ricerca e degli orizzonti limitati tentano di fare dei progressi sui problemi difficili prendendo qualunque scorciatoia *essi* intravedano, pur essendo molto critici sulle scorciatoie degli altri. Nessuno può abbracciare con chiarezza tutti i problemi e i relativi dettagli, compreso il sottoscritto, e tutti devono limitarsi a borbottare, tirare a indovinare e fare ipotesi approssimative su gran parte dei problemi.

Per esempio, uno dei rischi professionali dei neuroscienziati sembra essere la tendenza a pensare alla coscienza come a un *capolinea*. (È come se ci si dimenticasse che il prodotto finale degli alberi di mele non sono le mele – ma altri alberi di mele.) Naturalmente i neuroscienziati hanno cominciato a permettersi di pensare alla coscienza solo molto recentemente e solo alcuni coraggiosi teorici hanno cominciato a parlare, ufficialmente, di quello che hanno pensato. Come il ricercatore sulla visione Bela Julesz ha recentemente detto scherzosamente, puoi davvero farla franca solo se hai i capelli bianchi – e un premio Nobel! Ecco, per esempio, un'azzardata ipotesi di Francis Crick e Christof Koch:

> Noi abbiamo suggerito che una delle funzioni della coscienza sia quella di presentare il risultato di varie computazioni sottostanti e che ciò implica un meccanismo dell'attenzione che collega temporaneamente i neuroni interessati sincronizzando i loro picchi a oscillazioni di 40 hz. (Crick e Koch, 1990, p. 272)

Così una funzione della coscienza è di *presentare il risultato di computazioni sottostanti* – ma a chi? Alla Regina? Crick e Koch non vanno avanti e non si pongono la Domanda Difficile: *e poi*

cosa succede? («E poi avviene un miracolo?»). Dopo aver fatto una breve sortita in ciò che considerano il magico cerchio della coscienza, la loro teoria si arresta. Non si confronta con i problemi che abbiamo visto nei capitoli 5-8, per esempio, sull'insidioso sentiero che va dalla (presunta) coscienza al comportamento, inclusi, specialmente, i resoconti introspettivi.

I modelli della mente prodotti dalla psicologia cognitiva e dall'IA, al contrario, non soffrono quasi mai di *questo* difetto (vedi, per es., Shallice, 1972, 1978; Johnson-Laird, 1983, 1988; Newell, 1990). Generalmente, in un modello del genere viene postulato uno «spazio di lavoro» o «memoria di lavoro» che sostituisce il Teatro Cartesiano, poi viene mostrato come i risultati delle computazioni che si svolgono lì alimentano altre computazioni che guidano il comportamento, forniscono le informazioni per i resoconti verbali, si ripiegano ricorsivamente su se stessi per fornire nuovi dati in ingresso alla memoria di lavoro, e così via. Ma una caratteristica tipica di questi modelli è che non dicono dove e come una memoria di lavoro potrebbe essere collocata nel cervello e sono così preoccupati con il *lavoro* che deve essere svolto in quello spazio di lavoro che non c'è tempo per il «divertimento» – nessun segno di quella sorta di *piacevolezza* della fenomenologia che sembra una caratteristica così importante della coscienza umana.

Curiosamente, allora, i neuroscienziati finiscono spesso per sembrare dualisti, poiché una volta che hanno «presentato» le cose nella coscienza sembrano passare il testimone alla Mente, mentre gli psicologi cognitivisti spesso finiscono per sembrare zombisti (automatisti?), poiché descrivono strutture sconosciute alla neuroanatomia e si sforzano di mostrare come tutto il lavoro possa essere svolto senza dover svegliare un Osservatore Interno.

Le apparenze sono però fuorvianti. Crick e Koch non sono dualisti (anche se sono, evidentemente, materialisti cartesiani) e gli psicologi cognitivisti non hanno *negato* l'esistenza della coscienza (anche se, la maggior parte del tempo, fanno del loro meglio per ignorarla). Inoltre questi approcci parziali non squalificano nessuno dei due tentativi. I neuroscienziati hanno ragione nell'insistere che non hai un modello davvero buono della coscienza fino a che non hai risolto il problema di dove inserirlo nel cervello, ma gli scienziati cognitivisti (gli «IAisti» e gli psicologi cognitivisti, per esempio) hanno ragione nell'insistere che non hai un modello davvero buono della coscienza fino a che non hai risolto il problema di quali funzioni esegue e come

le esegue – meccanicamente, senza il beneficio della Mente. Come Philip Johnson-Laird si è espresso: «Qualunque teoria scientifica della mente deve trattarla come un automa» (Johnson-Laird, 1983, p. 477). L'orizzonte limitato di ciascun tentativo preso per se stesso mostra solo la necessità di un altro tentativo – quello in cui siamo impegnati – che prova a mettere insieme il più possibile i punti di forza di ciascuno.

2. *Orientiamoci con lo schizzo in miniatura*

Il compito principale di questo libro è filosofico: mostrare come una teoria veramente esplicativa della coscienza *possa* essere costruita a partire dalle considerazioni che stiamo svolgendo, non costruire – e confermare – una tale teoria in tutti i suoi dettagli. Ma la mia teoria sarebbe stata inconcepibile (almeno per me) se non avesse attinto copiosamente dal lavoro empirico di varie discipline che ha aperto (almeno a me) nuove prospettive di pensiero. (Una collezione particolarmente ricca di scoperte empiriche e nuove idee sulla coscienza è Marcel e Biasiach, 1988.) Per chi voglia coprirsi di gloria, questo è il momento giusto per tuffarsi nelle ricerche sulla mente. L'aria è così densa di nuove scoperte, nuovi modelli, sorprendenti risultati sperimentali – e pressoché altrettante millantate «dimostrazioni» e premature refutazioni. Oggi, le frontiere della ricerca sulla mente sono così aperte che non c'è nessun criterio assodato per stabilire quali siano le domande giuste e i metodi giusti. Con così tanti frammenti di teorie e congetture poco difendibili, è una buona idea posporre le nostre esigenze dimostrative e cercare invece delle basi più o meno indipendenti, seppur non concludenti, che tendono a convergere in difesa di una singola ipotesi. Non dobbiamo farci trasportare dal nostro entusiasmo, comunque. Talvolta ciò che sembra un fumo sufficiente a garantire l'esistenza di un fuoco robusto è in realtà solo una nuvola di polvere provocata da un carro di passaggio.

Nel suo libro *A Cognitive Theory of Consciousness* (1988), lo psicologo Bernard Baars riassume così ciò che gli sembra un «accordo crescente» sulla natura della coscienza: essa è realizzata da «una società distribuita di specialisti dotata di una memoria di lavoro, o *luogo di lavoro globale*, i cui contenuti possano essere inviati all'intero sistema» (p. 42). Come egli nota, un gruppo variegato di teorici, malgrado le enormi differenze di prospettive, formazione e aspirazioni, sta gravitando

verso questa posizione comune sul modo in cui la coscienza dovrebbe risiedere nel cervello. È una versione di questo accordo crescente che io sto cautamente presentando, ignorando alcune caratteristiche ed enfatizzandone altre – caratteristiche che ritengo siano state trascurate o sottovalutate, ma che ritengo particolarmente cruciali per aprirsi un varco nei misteri concettuali che ancora permangono.

Per orientare la mia teoria rispetto ad alcune delle montagne di lavoro da cui ho attinto, torniamo al mio schizzo in miniatura e ripercorriamolo un tema alla volta, rilevando i parallelismi e indicando le fonti e i disaccordi.

Non c'è un unico e definitivo «flusso di coscienza», perché non c'è un Quartier Generale centrale, un Teatro Cartesiano dove «tutto converge» per essere attentamente scrutinato da un Autore Centrale...

Sebbene tutti siano d'accordo nel dire che non esiste un unico punto del genere nel cervello, che richiama alla mente la ghiandola pineale di Cartesio, non da tutti sono state riconosciute le implicazioni di questo fatto, e talvolta sono state grossolanamente sorvolate. Per esempio, incaute formulazioni del «problema del collegamento» nell'attuale ricerca neuroscientifica spesso presuppongono che ci debba essere un unico spazio rappresentazionale nel cervello (più piccolo dell'intero cervello) dove i risultati di tutte le varie discriminazioni vengono fatti collimare – il sonoro viene sincronizzato con la pellicola, le forme vengono colorate, le parti vuote vengono riempite. Esistono alcune formulazioni scrupolose del problema (dei problemi) del collegamento che evitano quest'errore, ma le minuzie sono spesso trascurate.

[...] Invece di un unico (per quanto ampio) flusso del genere, ci sono canali multipli in cui vari circuiti specializzati tentano, in un pandemonio parallelo, di fare varie cose, creando man mano delle Molteplici Versioni. La maggior parte di queste frammentarie versioni di «narrazioni» giocano dei ruoli effimeri nella modulazione dell'attività in corso. [...]

Nell'IA, l'importanza di sequenze simili a narrazioni è stata a lungo sottolineata da Roger Schank, prima nel suo lavoro sugli *scripts* (1977, con Abelson) e più recentemente (1991) nel suo lavoro sul ruolo del raccontare storie nella comprensione. Da prospettive molto differenti, sempre all'interno dell'IA,

Patrick Hayes (1979), Marvin Minsky (1975), John Anderson (1983) e Erik Sandeval (1991) – e altri – hanno sostenuto l'importanza di strutture di dati che non siano solo sequenze di «istantanee» (con il conseguente problema di reidentificare i particolari nei *frames* successivi), ma siano invece specificamente progettate in un modo o nell'altro per rappresentare direttamente sequenze temporali e tipi di sequenze. In filosofia, Gareth Evans (1982) aveva cominciato a sviluppare alcune idee analoghe prima della sua morte prematura. In neurobiologia, questi frammenti narrativi sono esplorati come *scenari* e altre sequenze da William Calvin (1987) nel suo approccio al cervello come Macchina Darwiniana. Gli antropologi da molto tempo asseriscono che i miti che ogni cultura trasmette ai suoi nuovi membri giocano un ruolo importante nel forgiare le loro menti (vedi, per es., Goody, 1977, e per una proposta di applicazione all'IA, Dennett, 1991b), ma non hanno fatto nessun tentativo di darne un modello computazionale o neuroanatomico.

[...] ma qualcuna viene promossa ad ulteriori ruoli funzionali, in rapida successione, dall'attività di una macchina virtuale nel cervello. La serialità di questa macchina (il suo carattere «neumanniano») non è una caratteristica progettuale «cablata rigidamente», ma piuttosto il risultato di una successione di coalizioni di questi specialisti [...]

Molti hanno messo in risalto l'incedere relativamente lento e goffo dell'attività mentale cosciente (per es., Baars, 1988, p. 120), e un'ipotesi latente da molto tempo suggerisce che ciò potrebbe essere dovuto al fatto che il cervello non è stato in realtà progettato – cablato rigidamente – per una tale attività. Da vari anni circola l'idea che la coscienza umana potrebbe essere l'attività di una specie di macchina virtuale e seriale implementata sull'hardware parallelo del cervello. Lo psicologo Stephen Kosslyn ha proposto una versione dell'idea della macchina virtuale seriale ad un congresso della Society for Philosophy and Psycology all'inizio degli anni Ottanta e, su per giù dalla stessa epoca, io sto saggiando differenti versioni di questa idea (per es., Dennett, 1982b), ma una presentazione precedente di un'idea molto simile – sebbene senza l'uso del termine «macchina virtuale» – è contenuta nel fecondo scritto di Paul Rozin, *The Evolution of Intelligence and Access to Cognitive Unconscious* (1976). Un altro psicologo, Julian Jaynes, nelle sue audaci e originali congetture ne *Il crollo della mente bicamerale e l'origine*

della coscienza (1976), ha sostenuto che la coscienza umana è una sovrapposizione molto recente, apportata dalla cultura, su una precedente architettura funzionale, un tema sviluppato in altri modi anche dal neuroscienziato Harry Jerison (1973). L'architettura neurale sottostante non è affatto una *tabula rasa* alla nascita, secondo quest'opinione, ma è tuttavia un mezzo in cui le strutture vengono costruite come una funzione delle interazioni del cervello con il mondo. Ed è a queste strutture costruite, piuttosto che a quelle innate, che bisogna ricorrere per spiegare il funzionamento cognitivo.

Gli specialisti fondamentali sono parte della nostra eredità animale. Non furono sviluppati per eseguire azioni specificamente umane, come leggere e scrivere, ma per evitare i predatori, schivare gli oggetti, riconoscere i volti, afferrare, scagliare, raccogliere le bacche e altri compiti essenziali [...]

Queste orde di specialisti sono proposte da varie teorie differenti, ma le loro dimensioni, i loro ruoli e la loro organizzazione è molto controversa. (Per una rassegna utile e veloce, vedi Allport, 1989, pp. 643-647.) I neuroanatomisti che studiano i vari cervelli animali dalle oloturie e le seppie ai gatti e le scimmie hanno identificato molte varietà di circuiti rigidamente cablati squisitamente progettati per eseguire compiti specifici. I biologi parlano di Meccanismi Innati di Rilascio (*Innate Releasing Mechanisms* o IRMs) e Strutture di Azioni Fisse (*Fixed Action Patterns* o FAPs), che possono essere accoppiati tra loro, e, in una lettera recente, il neuropsicologo Lynn Waterhouse mi ha descritto in modo calzante le menti degli animali come composte di vari «collage di IRMs-FAPs». Sono proprio queste menti animali problematicamente assemblate in tali collage che Rozin (assieme ad altri) presuppone come base per l'evoluzione di menti con obiettivi più generali, che sfruttano questi meccanismi preesistenti per nuovi scopi. Lo psicologo percettivo V.S. Ramachandran (1991) nota che «c'è un vantaggio reale che si presenta con i sistemi multipli: ti permette la tolleranza per le immagini sfuocate del tipo che incontri nel mondo reale. La mia analogia favorita per illustrare alcune di queste idee è quella dei due ubriachi: nessuno dei due può camminare senza sostegno, ma appoggiandosi l'uno sull'altro riescono a barcollare verso la loro meta».

Il neuropsicologo Michael Gazzaniga ha indicato una massa di dati che sorgono da deficienze neurologiche (inclusi

quelli famosi ma spesso fraintesi relativi ai pazienti con cervello diviso) come prove che la mente può essere considerata una coalizione o fascio di agenzie semi-indipendenti (Gazzaniga e Ledoux, 1978; Gazzaniga, 1985); e, da un'angolazione differente, il filosofo della psicologia Jerry Fodor (1983) ha sostenuto che parti molto ampie della mente umana sono composte di *moduli*: sistemi specializzati, rigidamente cablati, «incapsulati», di analisi dei dati in ingresso (e di generazione di dati in uscita – sebbene non abbia avuto molto da dire su questo).

Fodor si concentra su moduli che apparterrebbero specificamente alla mente umana – moduli per l'acquisizione del linguaggio e l'analisi delle proposizioni, in particolare – e poiché egli ignora completamente il problema di una loro genesi da probabili antenati nelle menti degli animali inferiori, contribuisce a creare l'impressione fuorviante che l'evoluzione abbia progettato i nuovi meccanismi per il linguaggio, tipici della specie umana, senza riutilizzare «pezzi usati». La sensazione che questi moduli siano un dono miracoloso che Madre Natura offre all'*Homo sapiens* è rafforzata dal modo ultra-intellettualistico in cui Fodor pensa che i moduli siano connessi al resto della mente. Secondo Fodor, essi non eseguono dei compiti completi nell'economia della mente (come controllare la coordinazione occhio-mano per raccogliere qualcosa), ma si arrestano improvvisamente ad una soglia interna, una linea nella mente che non possono oltrepassare. C'è un'arena centrale in cui si determina una razionale «fissazione delle credenze», sostiene Fodor, in cui i moduli depositano servilmente i loro beni, consegnandoli a processi *non* modulari («globali, isotropici»).

I moduli di Fodor farebbero la felicità di un burocrate: la definizione dei loro compiti è scolpita nel marmo; essi non possono essere utilizzati per svolgere nuovi o molteplici ruoli; e sono «cognitivamente impenetrabili» – il che significa che la loro attività non può essere modulata, o anche solo interrotta, da cambiamenti negli stati informazionali «globali» del resto del sistema. Secondo Fodor, tutte le attività realmente intelligenti della cognizione sono non modulari. Pensare a cosa fare dopo, ragionare su situazioni ipotetiche, ristrutturare creativamente i propri materiali, rivedere le proprie opinioni sul mondo – tutte queste attività sono eseguite da un'apparecchiatura misteriosa e centrale. Inoltre, Fodor sostiene (con curiosa soddisfazione) che nessun ramo della scienza cognitiva – inclusa la filosofia – ha la minima idea su come funziona questa apparecchiatura centrale!

Si sa molto sulle trasformazioni delle rappresentazioni che servono per mettere le informazioni in una forma appropriata all'elaborazione centrale; non si sa praticamente nulla su cosa avviene dopo che le informazioni arrivano lì. Lo spettro è stato ricacciato nel più profondo della macchina, ma non è stato ancora esorcizzato. (Fodor, 1983, p. 27)

Dando a questa apparecchiatura centrale tanto da fare, e tanto potere non modulare con cui farlo, Fodor trasforma i suoi moduli in agenti poco plausibili, agenti la cui esistenza acquista un senso solo all'interno di una squadra con un Agente Boss che ha una preoccupante autorità (Dennett, 1984b). Poiché nel descrivere i suoi moduli Fodor è stato attento soprattutto a differenziare la loro meccanicità finita, comprensibile, e non mentale dagli illimitati e inesplicabili poteri del centro non modulare, molti di quelli che sarebbero stati altrimenti favorevoli ad almeno gran parte della sua caratterizzazione dei moduli hanno avuto la tendenza a liquidarli come le fantasie di un cripto-cartesiano.

Molti di quegli stessi teorici si sono mostrati tra il tiepido e l'ostile nei confronti degli *Agenti* di Marvin Minsky, che formano *La società della mente* (1985). Gli Agenti di Minsky sono homunculi di tutte le dimensioni, dal gigantesco specialista con talenti così elaborati da assomigliare ai moduli fodoriani, fino al piccolo agente delle dimensioni di un meme (polinemi, micronemi, agenti censori, agenti soppressori, ed altri ancora). Sembra troppo facile, pensano gli scettici, postulare una banda di agenti fatti su misura per quell'incarico ogni volta che c'è qualcosa da fare – una mossa teoretica che ha tutti i vantaggi che il furto ha rispetto all'onesto lavoro, per riadattare una famosa espressione di Bertrand Russell.

Gli homunculi – i folletti, gli agenti – costituiscono la moneta corrente nell'Intelligenza Artificiale, e più in generale nell'informatica. Chiunque storca le ciglia appena sente nominare gli homunculi, semplicemente non ha capito quanto questo concetto possa essere neutrale, e quanto possa essere ampiamente applicato. Postulare una banda di homunculi sarebbe sicuramente una mossa inutile come immagina lo scettico, se non fosse che, nelle teorie homuncolari, ciò che è davvero importante sono le affermazioni sul modo in cui gli homunculi postulati interagiscono, si sviluppano, formano coalizioni o gerarchie, e così via. E in questo le teorie possono essere davvero molto differenti. Le teorie burocratiche, come abbiamo

visto nel capitolo 8, organizzano gli homunculi in gerarchie predefinite. Non ci sono homunculi lavativi o disubbidienti, e la competizione tra gli homunculi è così strettamente regolata come quella di un torneo sportivo. Le teorie in stile pandemonio, al contrario, postulano molti sforzi ripetuti, sprechi di energia, interferenze, periodi caotici e perdigiorno senza un lavoro ben determinato. Chiamare homunculi (o folletti o agenti) le unità di queste teorie molto differenti non è più informativo che chiamarle semplicemente... unità. Sono solamente unità con competenze particolari e circoscritte, ed ogni teoria, dalla più rigorosamente neuroanatomica alla più astrattamente artificiale, non fa altro che postulare tali unità per poi descrivere come possano essere ottenute delle funzioni più complesse organizzando le unità che eseguono funzioni più semplici. In effetti *tutte* le varietà di funzionalismo possono essere considerate come funzionalismo «omuncolare» in un modo o nell'altro.

Mi sono divertito a notare un curioso eufemismo che ha recentemente guadagnato i favori dei neuroscienziati. I neuroanatomisti hanno fatto progressi enormi nel mappare la corteccia, che si è rivelata essere meravigliosamente organizzata in colonne specializzate di neuroni interagenti (il neuroscienziato Vernon Mountcastle, 1978, le chiama «moduli unitari») organizzate in ulteriori strutture più grandi come le «mappe retinotopiche» (in cui la configurazione spaziale dell'eccitazione sulla retina degli occhi viene conservata), che a loro volta giocano ruoli – ancora non ben compresi – in organizzazioni ancora più grandi di neuroni. Una volta, i neuroscienziati avevano l'abitudine di parlare di quello che *segnalano* questi vari tratti o gruppi di neuroni nella corteccia; stavano pensando a queste unità come a homunculi il cui «compito» è sempre quello di «inviare un messaggio con un determinato contenuto». I recenti sviluppi hanno suggerito che questi tratti possano eseguire delle funzioni molto più complesse e variegate, e così ora viene considerato fuorviante dire che essi (semplicemente) segnalano questo o quello. Come possiamo, allora, esprimere queste scoperte sofisticate sulle specifiche condizioni in cui questi tratti sono attivi? Noi diciamo che questo tratto «si occupa del» colore, mentre quello «si occupa del» movimento. Ma questo uso non è affatto un ridicolo antropomorfismo o un «errore omuncolare» del tipo di quelli che si incontrano ovunque nell'IA! Grazie a Dio, questo è solamente un modo intelligente che sobri ricercatori hanno trovato per parlare delle

competenze dei tratti nervosi in modo suggestivo ma senza entrare in modo indebito nei particolari! In fondo, se non è zuppa è pan bagnato.

Gli Agenti di Minsky si distinguono principalmente perché, a differenza della maggior parte degli altri tipi di homunculi postulati, hanno una storia e una genealogia. La loro esistenza non viene solo postulata; viene fatto vedere come deve essersi sviluppata a partire da qualcosa la cui esistenza precedente non è completamente misteriosa; e Minsky dà molti suggerimenti su come tali sviluppi devono avvenire. Se dà fastidio che egli ancora non si pronunci sulla esatta localizzazione degli agenti nel cervello e sui neuroni che dovrebbero costituirli, si tenga presente che il suo scopo era di indagare i requisiti più generali sullo sviluppo delle funzioni senza essere eccessivamente dettagliato. Come egli dice, descrivendo la sua precedente teoria dei *frames* (di cui la Società della Mente è un risultato), «Se la teoria fosse stata ancora più vaga, sarebbe stata ignorata, ma se fosse stata descritta più dettagliatamente, altri scienziati l'avrebbero forse "collaudata", invece di dare il loro contributo con le loro idee» (1985, p. 506). Alcuni scienziati rimangono insensibili a questa giustificazione. Essi sono interessati solo a quelle teorie che producono previsioni immediatamente sottoponibili a prova. Questa potrebbe anche essere una giusta ostinazione, se non fosse per il fatto che tutte le teorie sottoponibili a prova fin qui architettate si sono dimostrate chiaramente false, e sarebbe ingenuo ritenere che le innovazioni richieste per comporre *nuove* teorie sottoponibili a prova possano sbocciare miracolosamente senza una buona dose di esplorazione fantasiosa del tipo in cui Minsky indulge. (Naturalmente, anch'io sto giocando allo stesso gioco.)

Ritorniamo al mio schizzo in miniatura:

Spesso [i folletti specializzati] sono opportunisticamente trasferiti in nuovi ruoli, ai quali sono più o meno portati a seconda dei loro talenti originali. Non ne risulta un manicomio solo perché le tendenze imposte a tutta questa attività sono esse stesse il prodotto di un progetto. Parte di questo progetto è innato ed è condiviso con altri animali. Ma viene integrato, e spesso sopraffatto, da microabitudini di pensiero che vengono sviluppate nell'individuo, in parte come risultati di auto-esplorazione propri dell'individuo stesso e in parte come doni predefiniti della cultura. Migliaia di memi, nella maggior parte trasportati dal linguaggio, ma anche da «imma-

gini» senza parole e altre strutture di dati, stabiliscono la loro residenza in un cervello individuale, forgiando le sue tendenze e trasformandolo quindi in una mente.

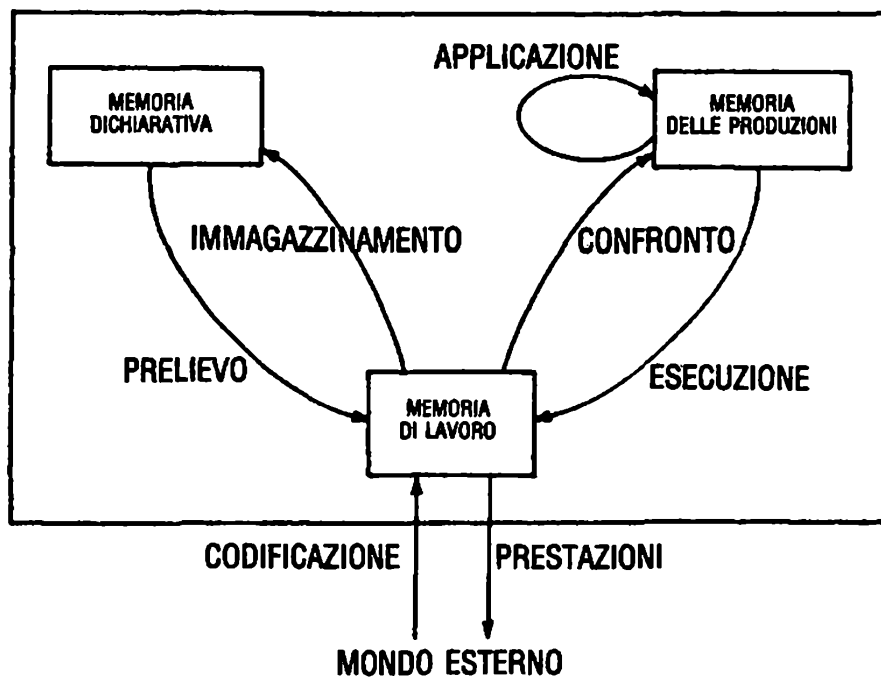
In questa parte della mia teoria ho evitato volutamente di essere più specifico su molte questioni importanti: come interagiscono in realtà questi homunculi per realizzare qualcosa? Quali sono le transazioni sottostanti che elaborano l'informazione, e che ragioni abbiamo per pensare che possano «funzionare»? Secondo il mio schizzo, la sequenza degli eventi è determinata (in modi che ho solo accennato) da «abitudini», e a parte alcune affermazioni negative nel capitolo 5 su ciò che *non* avviene, io finora non sono stato affatto preciso sulla struttura dei processi grazie ai quali degli elementi delle Molteplici Versioni vengono conservati, alcuni dei quali finiscono per generare l'eterofenomenologia come risultato di un qualche sondaggio. Per vedere a cosa si riduce la questione e quali potrebbero essere le risposte alternative, dovremmo osservare brevemente qualche modello più esplicito di pensiero sequenziale.

3. *E poi cosa succede?*

Nel capitolo 7 abbiamo visto come l'architettura di von Neumann sia un distillato del processo seriale di calcolo deliberato. Turing e von Neumann isolarono un aspetto particolare della corrente che scorre nel flusso della coscienza e poi lo hanno radicalmente idealizzato per meccanizzarlo. C'è la nota strozzatura di von Neumann, costituita da un unico registro per i risultati e un unico registro per le istruzioni. I programmi sono semplicemente delle liste ordinate di istruzioni tratte da un piccolo insieme di istruzioni primitive che la macchina può eseguire perché sono cablate rigidamente in essa. Un processo fisso, il ciclo di prelievo, recupera le istruzioni dalla coda nella memoria, una per volta, andando sempre a prendere l'istruzione successiva nella lista, a meno che una precedente istruzione non lo dirotti verso un'altra parte della lista.

Quando i ricercatori dell'IA incominciarono a implementare modelli più realistici delle operazioni cognitive basandosi sull'architettura di von Neumann, vi apportarono profonde modifiche. Ampliarono la strozzatura esageratamente angusta di von Neumann, trasformandola in un «luogo di lavoro» o

«memoria di lavoro» in qualche modo più capace. Progettarono anche delle operazioni più sofisticate che servissero da funzioni psicologiche primitive e sostituirono l'istruzione del rigido ciclo di prelievo della macchina di von Neumann con maniere più flessibili di richiamare ed eseguire le istruzioni. Lo spazio di lavoro è diventato, in qualche caso, una «lavagna» (Reddy e altri, 1973; Hayes-Roth, 1985), su cui vari folletti possono scrivere vari messaggi che gli altri possono leggere, il che provoca a sua volta un altro giro di scrittura e lettura. L'architettura di von Neumann, con il suo rigido ciclo di istruzioni, era ancora lì, nello sfondo, a realizzare l'implementazione, ma non giocava più un ruolo nel modello. Ciò che accadeva poi nel modello era dettato dai risultati di ondate competitive di scrittura e lettura di messaggi sulla lavagna. Una specie correlata di discendenti dell'architettura di von Neumann è quella composta dai vari *sistemi di produzione* (Newell, 1973) che sottostanno a modelli come ACT* (leggi: *act star*, «ACT-stella») di John Anderson (1983) e Soar di Rosenbloom, Laird e Newell (1987); (vedi anche Newell, 1990).

Ci si può fare una buona idea dell'architettura sottostante ad un sistema di produzione da questa semplice illustrazione di ACT*.

9.1

La memoria di lavoro è dove avvengono tutte le azioni. Tutte le azioni elementari sono chiamate *produzioni*. Le produzioni sono in definitiva solo dei meccanismi di riconoscimento di configurazioni che scattano ogni qualvolta rilevino la *loro*

configurazione specifica. Sono, cioè, operatori SE-ALLORA che si aggirano nella memoria di lavoro osservando i vari contenuti attualmente presenti, aspettando che i loro enunciati SE siano soddisfatti, per dare il permesso agli ALLORA di svolgere le loro imprese, quali che siano (in un tipico sistema di produzione: deporre un nuovo dato nella memoria di lavoro, per l'ulteriore scrutinio da parte delle produzioni).

Tutti i calcolatori hanno istruzioni primitive SE-ALLORA, esse sono gli «organi di senso» che rendono possibile reagire in modo differenziato ai dati che sopraggiungono dall'esterno o sono richiamati dalla memoria. Questa capacità di una *deviazione condizionale* è un ingrediente essenziale del potere dei calcolatori, indipendentemente dalla loro architettura. I SE-ALLORA originali erano le semplici e chiare istruzioni di stato di macchina ideate da Turing: SE vedi zero, ALLORA sostituiscilo con uno, spostati di uno spazio a sinistra e mettiti nello stato *n*. Paragoniamo queste semplici istruzioni con il tipo molto più complesso che potrebbero essere date ad una sentinella umana abile e ben addestrata: SE vedi qualcosa che ti sembra sospetta E dopo un'indagine sommaria rimane tale OPPURE sopraggiungono altri dubbi, ALLORA dai l'allarme. Possiamo costruire questo sofisticato monitoraggio a partire dai semplici e meccanici SE-ALLORA? Le *produzioni* sono dei sensori di livello intermedio dai quali si possono costruire organi di senso più complessi, e poi intere architetture cognitive. Le produzioni si possono servire di enunciati SE complessi e confusi; le configurazioni che riconoscono non devono essere semplici come i codici a barre identificabili dai registratori di cassa, ma possono assomigliare di più a configurazioni come quella discriminata dalla sentinella (vedi a questo proposito la discussione in Anderson, 1983, pp. 35-44). E a differenza dei SE-ALLORA di una macchina di Turing, che è sempre in un unico stato di macchina alla volta (esaminando sempre un unico SE-ALLORA tra quelli che ha in memoria, prima di passare al prossimo dato), i SE-ALLORA di un sistema di produzione aspettano *in massa*, in parallelo (simulato), cosicché a ogni «istante» più di una produzione può avere la sua condizione soddisfatta e essere pronta all'azione.

È qui che le cose diventano interessanti: come fa un sistema del genere a *risolvere i conflitti*? Quando viene soddisfatta più di una produzione, c'è sempre la possibilità che due (o più) di loro spingano verso direzioni incompatibili. I sistemi paralleli possono tollerare una grossa quantità di obiettivi conflittuali, ma in un sistema che deve riscuotere successo nel mondo non tutto

può avvenire alla stesso tempo; talvolta qualcosa deve cedere. Un aspetto cruciale nella differenziazione dei sistemi è il modo in cui trattano la risoluzione dei conflitti. Infatti, poiché la maggior parte dei dettagli psicologici o biologici più interessanti si trovano nelle diversità che emergono a questo livello, è meglio considerare l'architettura di un sistema di produzione come un semplice mezzo sottostante con cui costruire i modelli. Ma tutti i sistemi di produzione condividono alcune presupposizioni basilari che forniscono un ponte per il nostro schizzo di teoria: essi hanno uno spazio di lavoro dove si svolge l'azione, dove molte produzioni (= folletti) possono cercare di svolgere le loro cose contemporaneamente; hanno una memoria più o meno inerte dove l'informazione innata o accumulata viene conservata. Poiché non tutto quello che «è conosciuto» dal sistema è disponibile contemporaneamente in questo spazio di lavoro, il problema di Platone di riuscire a richiamare l'uccello giusto al momento giusto è il principale compito logistico da affrontare. E, cosa molto importante per il nostro attuale e proficuo modo di vedere, i teorici *hanno progettato realmente* dei meccanismi da candidare come risposte alla Domanda Difficile: *e poi cosa succede?*

Per esempio, in ACT*, ci sono cinque princìpi per la risoluzione dei conflitti:

1. *Grado di accordo*: se l'enunciato SE di una produzione si accorda meglio di quello di un'altra, essa ha la priorità.
2. *Forza della produzione*: le produzioni che hanno avuto recentemente successo hanno una «forza» superiore, e quindi la priorità su quelle con minore forza.
3. *Refrattarietà ai dati*: la stessa produzione non può accoppiarsi allo stesso dato per più di una volta (questo per impedire cicli infiniti e simili fossilizzazioni, seppur men gravi).
4. *Specificità*: quando due produzioni si accordano allo stesso dato, la produzione con l'enunciato SE più specifico vince.
5. *Dominanza dell'obiettivo*: tra le cose che le produzioni depositano nella memoria di lavoro ci sono gli obiettivi. Nella memoria di lavoro di ACT* ci può essere solo un obiettivo attualmente attivo, e ogni produzione con un risultato che si accorda all'obiettivo attivo ha la priorità.

Questi sono tutti principi plausibili per la risoluzione dei conflitti, sono sensati sia psicologicamente sia teleologicamente (per una discussione dettagliata, vedi Anderson, 1983, cap. 4).

Ma forse sono troppo sensati. Cioè, Anderson si è messo a progettare assennatamente il sistema di risoluzione dei conflitti di ACT* sfruttando la sua conoscenza del tipo specifico di problemi che sorgono nelle situazioni di risoluzione dei conflitti e dei modi efficaci di affrontarli. Ha sostanzialmente cablato rigidamente questa sofisticata conoscenza nel sistema, un dono innato dell'evoluzione.

È interessante considerare per contrasto l'architettura Soar di Rosenbloom, Laird e Newell (1987). Anch'essa, come ogni architettura parallela, incontra delle *impasse* – occasioni in cui serve una risoluzione di un conflitto a causa o di produzioni contraddittorie o di produzioni che «non scattano» – ma le considera come doni, non problemi. Le *impasse* sono le opportunità basilari per costruire all'interno del sistema. I conflitti non sono trattati automaticamente servendosi di un insieme prefissato di principi per la risoluzione dei conflitti (un autoritario homunculus vigile urbano che regola il traffico già sul posto) ma piuttosto in modo *non* automatico. Un'*impasse* crea un *nuovo* «spazio per problemi» (una sorta di spazio topico di lavoro) in cui il problema da risolvere è proprio l'impasse stessa. Ciò può generare un altro spazio per problemi meta-meta-traffico, e così via – *potenzialmente* per sempre. Ma in pratica (almeno nei domini studiati finora), dopo avere accumulato spazi per problemi uno sopra l'altro, il problema più in alto trova una soluzione, il che risolve rapidamente il problema immediatamente più basso, e così via, dissolvendo la preoccupante proliferazione di spazi dopo aver effettuato una esplorazione non banale attraverso lo spazio logico delle possibilità. Il sistema, inoltre, si modifica perché «ingloba» la scoperta così difficilmente ottenuta in nuove produzioni, di modo che, quando si verificheranno problemi simili, ci sarà una produzione nuova di zecca a portata di mano per risolverli rapidamente, saranno considerati banali problemi già risolti in passato.

Ho fatto rapidamente cenno a questi dettagli non per sostenere una superiorità di Soar su ACT*, ma solo per dare un'idea del tipo di problemi che possono essere indagati, seriamente, da modelli del genere. La mia impressione è che, per varie ragioni di cui non abbiamo bisogno di interessarci ora, il mezzo sottostante dei sistemi di produzione è *ancora* troppo idealizzato e semplificato nei suoi princìpi limitativi, ma lo sviluppo che ci ha portato dalla macchina di von Neumann ai sistemi di produzione ci indica altre architetture ulteriori, con strutture sempre più ingegnose, e il modo migliore per verificare la loro forza e le

loro debolezze è di costruirle e farle girare. Questo è il modo con il quale è possibile trasformare ciò che è appena abbozzato e vago all'interno di teorie come la mia in modelli dignitosi con dettagli che possano essere empiricamente messi alla prova.

Quando si prendono le varie affermazioni sui meccanismi della coscienza che ho fatto negli ultimi quattro capitoli e si incomincia a sovrapporli a modelli di sistemi cognitivi come questi, sorge una serie di questioni, a cui però io non proverò a rispondere ora. Poiché lascio tutte queste questioni irrisolte, il mio schizzo rimane quello che è – uno schizzo che potrebbe adattarsi a un'intera famiglia di teorie molto differenti. In questa occasione io non devo spingermi oltre, giacché i problemi filosofici della coscienza ruotano attorno alla possibilità o meno che una *qualsiasi* teoria del genere spieghi la coscienza, così sarebbe prematuro affidare le nostre speranze a una versione molto particolare che potrebbe rivelarsi seriamente difettosa. (Nell'Appendice B mi spingerò su qualche ramo empirico, comunque, per quelli che vogliono che una teoria abbia delle implicazioni sottoponibili a prova sin dall'inizio.)

Non sono solo le teorie dei filosofi ad aver bisogno di essere messe a punto grazie alla costruzione di modelli a questo livello; le teorie dei neuroscienziati stanno sulla stessa barca. Per esempio, nell'elaborata teoria dei circuiti «rientranti» nel cervello, Gerald Edelman (1989) fa molte asserzioni sul modo in cui tali rientranti possano effettuare le discriminazioni, costruire le strutture di memoria, coordinare i passi sequenziali della risoluzione dei problemi e in generale eseguire le attività di una mente umana, ma a dispetto della ricchezza di dettagli neuroanatomici e delle asserzioni entusiastiche e spesso plausibili di Edelman, non sapremo che cosa possano fare i suoi rientranti – non sapremo se i rientranti siano il modo *giusto* di concepire la neuroanatomia funzionale – fino a che essi non siano inseriti in un'intera architettura cognitiva al livello di dettaglio di ACT* o Soar e messi sulle loro gambe.[1]

A un livello più dettagliato di costruzione di modelli, troviamo il problema non risolto di far vedere come le produzioni stesse (o comunque vogliamo chiamare i folletti riconoscitori di configurazioni) siano implementate nel cervello. Baars (1988) chiama i suoi specialisti «mattoni» con cui costruire, e opta per lasciare ad un altro giorno o un'altra disciplina gli ulteriori dettagli che sorgono con la fabbricazione dei mattoni; ma, come molti hanno notato, è allettante supporre che gli

specialisti stessi, a svariati livelli di aggregazione, siano fatti, in un modo o nell'altro, di tessuto *connessionista.*

Il connessionismo (o PDP, Parallel Distributed Processing, per elaborazione distribuita in parallelo) è uno sviluppo piuttosto recente nell'IA che promette di portare i modelli cognitivi più vicino ai modelli neurali, poiché gli elementi che costituiscono i *suoi* mattoni sono nodi in reti parallele che sono connesse in modi che sembrano *piuttosto* simili alle reti neurali nel cervello. Paragonare l'IA connessionista alla «IA di buon vecchio stile» (Haugeland, 1985) e a vari modelli progettati nelle neuroscienze è diventato la principale occupazione nel mondo accademico (vedi, per es., Graubard, 1988; Bechtel e Abrahamson, 1991; Ramsey, Stich e Rumelhart, 1991). Ciò non sorprende, poiché il connessionismo apre la prima strada di collegamento lontanamente plausibile attraverso la terra incognita che si estende tra le scienze della mente e le scienze del cervello. Ma quasi nessuna delle controversie che circondano «il modo appropriato di trattare il connessionismo» (Smolensky, 1988) influenzano i nostri progetti attuali. *Naturalmente* ci dovrà essere un livello (o livelli) della teoria simile a quello dei modelli connessionisti e medierà tra i livelli della teoria più ovviamente neuroanatomici e i livelli della teoria più ovviamente psicologici o cognitivi. La questione è quale particolare idea connessionista sarà parte di quella soluzione e quale sarà abbandonata strada facendo. Fino a quando ciò non sarà stabilito, i pensatori tendono ad usare l'arena del dibattito connessionista come un amplificatore per i loro slogan favoriti, e sebbene io desideri dire la mia in questi dibattiti come tutti gli altri (Dennett, 1987b, 1988b, 1989, 1990c, 1991b, c, d), mi morderò la lingua in questa occasione e continuerò a occuparmi del nostro compito principale, che è di vedere come una teoria della *coscienza* possa emergere da ciò dopo che la polvere si sia posata, qualunque sia il modo in cui ciò avvenga.

Si noti che quello che è avvenuto nella progressione dall'architettura di von Neumann a tali architetture virtuali come i sistemi di produzione e (a un livello più fine di analisi) ai sistemi connessionisti è stato quello che può essere chiamato uno spostamento nell'equilibrio del potere. Programmi fissi e progettati in anticipo, che corrono su binari con pochi scambi regolati dai dati, sono stati sostituiti da sistemi flessibili – in effetti, volatili – il cui comportamento successivo dipende molto di più dalle complesse interazioni tra ciò che il sistema sta attualmente incontrando e ciò che ha incontrato nel passato. Come Newell, Rosenbloom e Laird (1989) dicono: «Così il

problema per un calcolatore ordinario è come essere interrotto, mentre per Soar e ACT* (e presumibilmente per la cognizione umana) è come rimanere focalizzato» (p. 119).

Visto tutto l'inchiostro che è stato versato su questo problema teorico, è importante sottolineare che si tratta di uno spostamento nell'*equilibrio* del potere e non di uno spostamento verso qualche modo «qualitativamente differente» di operare. Nel cuore del più volatile sistema di riconoscimento delle configurazioni («connessionista» o no) giace un motore di von Neumann, che scoppiettando calcola una funzione computabile. Fin dalla nascita dei calcolatori, i critici dell'Intelligenza Artificiale si sono accaniti contro la rigidità, la meccanicità, la *programmaticità* dei calcolatori e i suoi difensori hanno ripetutamente sostenuto che è solo una questione di grado di complessità – che sui calcolatori potrebbero essere creati sistemi indefinitamente non-rigidi, vaghi, olistici, organici. Con lo sviluppo dell'IA, sono comparsi proprio dei sistemi del genere, così i critici devono ora decidere se prendere o lasciare. Debbono dichiarare che i sistemi connessionisti – per esempio – siano proprio come essi avevano sempre pensato che fossero le menti oppure debbono alzare la posta al buio e insistere nel dire che nemmeno un sistema connessionista è abbastanza «olistico» per loro o abbastanza «intuitivo» o abbastanza ... (inserisci il tuo slogan favorito). Due tra i più noti critici dell'IA, i filosofi di Barkeley Hubert Dreyfus e John Searle, si sono divisi su questo punto; Dreyfus ha giurato fedeltà al connessionismo (Dreyfus e Dreyfus, 1988), mentre Searle ha rialzato la posta, insistendo che nessun calcolatore connessionista potrebbe esibire *reali* caratteri mentali (1990a, 1990b).

Gli scettici «in linea di principio» possono essere in ritirata, ma enormi problemi aspettano ancora gli unificatori. Il più grande, secondo me, ha una diretta rilevanza sulla nostra teoria della coscienza. C'è un accordo diffuso nelle scienze cognitive sul fatto, illustrabile da dozzine di disegni come quello di figura 9.1, che *lì* abbiamo la memoria a lungo termine (la colombaia di Platone) e *qui* abbiamo lo spazio o la memoria di lavoro, dove il pensiero effettivamente si verifica.[2] Eppure non ci sono due posti nel cervello che ospitano queste due apparecchiature distinte. L'unico luogo nel cervello che sia una dimora plausibile per ciascuna di queste due separate funzioni è l'intera corteccia – non due spazi uno accanto all'altro ma un unico grande spazio. Come Baars dice, riassumendo l'accordo crescente, c'è uno spazio di lavoro *globale*. È globale non solo nel

senso funzionale (detto crudamente, è un «luogo» dove ogni cosa può entrare in contatto con ogni altra cosa), ma anche nel senso anatomico (è diffuso su tutta la corteccia e senza dubbio implica anche altre regioni del cervello). Ciò significa, allora, che lo spazio di lavoro deve avvalersi degli stessi tratti e reti neurali che apparentemente giocano un ruolo fondamentale nella memoria a lungo termine: «l'immagazzinamento» dei cambiamenti progettuali prodotti dall'esplorazione individuale.

Supponi di imparare a fare la focaccia di mais o di imparare il significato della parola «fenotipico». In qualche modo, la corteccia dovrà essere un mezzo in cui alcuni schemi stabili di connessione possano ancorare in modo quasi permanente questi emendamenti progettuali del cervello con cui sei nato. Se, per esempio, ti ricordi improvvisamente dell'appuntamento con il dentista, ciò ti toglie tutto il piacere che provavi dalla musica che stavi ascoltando. In qualche modo, la corteccia dovrà essere un mezzo in cui alcuni schemi *in*stabili di connessione possano alterare rapidamente questi contenuti momentanei dell'intero «spazio» – senza, naturalmente, cancellare in questo processo la memoria a lungo termine. Come possono questi due tipi molto differenti di «rappresentazione» coesistere nello stesso mezzo allo stesso tempo? Nei modelli puramente cognitivi, i compiti possono essere collocati in due riquadri separati di un diagramma, ma se dobbiamo sovrapporli alla trama unica del tessuto neurale, il semplice problema dell'impacchettamento è quello che ci preoccupa di meno.

È possibile supporre che due sistemi funzionalmente distinti si compenetrino a vicenda (come la rete telefonica e la rete autostradale si estendono sullo stesso territorio) – ma non è questo il punto. Il problema più profondo sta immediatamente dietro la superficie di un'ipotesi che abbiamo fatto. Abbiamo ipotizzato che i singoli folletti specialisti reclutano in qualche modo altri compagni per imprese su scala più ampia. Se questo riguardasse solamente il problema di invitare le nuove reclute a esercitare i loro talenti *specialistici* nella causa comune, noi avremmo già alcuni modelli di un processo simile – per esempio, ACT*, Soar e lo Spazio di Lavoro Globale di Baars – con vari gradi di plausibili dettagli. Ma cosa succederebbe se gli specialisti fossero talvolta reclutati anche come *generici*, per contribuire a funzioni in cui i loro talenti specializzati non giocano alcun ruolo discernibile? Questa è un'idea allettante, per varie ragioni (vedi, per es., Kinsbourne e Hicks, 1978), ma, per quanto ne sappia, non abbiamo ancora alcun modello

computazionale di come tali elementi a duplice funzione possano operare.

Ecco la difficoltà. Comunemente si suppone che gli specialisti nel cervello debbano in qualche modo ricevere la loro identità funzionale dalla loro effettiva posizione in una rete di connessioni più o meno fisse. Per esempio, sembra che l'unico genere di fatti che potrebbe spiegare perché un particolare tratto neurale «si occupa del» colore sia quello che riguarda fatti sulle sue particolari connessioni, anche indirette, con le cellule coniche nella retina che sono altamente sensibili alle differenti frequenze luminose. Una volta stabilita una tale identità funzionale, queste connessioni potrebbero essere recise (come accade talvolta in chi diventa cieco da adulto) senza una perdita (totale) della capacità degli specialisti di rappresentare (o in qualche altra maniera di «occuparsi di») il colore, ma senza tali originarie connessioni causali, è difficile capire che cosa potrebbe dare agli specialisti un ruolo connesso a un contenuto specifico.[3] Sembra allora che la corteccia sia (ampiamente) composta da elementi i cui poteri rappresentazionali più o meno fissi sono il risultato della loro posizione funzionale nella rete generale. Essi rappresentano nello stesso modo in cui i membri della Camera dei Deputati rappresentano le circoscrizioni: trasportando informazioni dalle fonti a cui sono specificamente collegati (la maggior parte delle chiamate telefoniche nei loro uffici della capitale provengono dalle loro circoscrizioni di origine, per esempio). Ora immagina i deputati seduti in un gruppo di posti contigui allo stadio e che, grazie a dei pannelli colorati tenuti al di sopra della loro testa, rappresentano l'importante messaggio «La velocità uccide!», le cui lettere gigantesche possono essere lette dall'altra parte dello stadio; sono, in breve, dei pixel viventi le cui relazioni con i loro elettori non giocano alcun ruolo nel contributo che danno alla rappresentazione di gruppo. Alcuni modelli del reclutamento corticale suggeriscono fortemente che *qualcosa di simile* a questo ruolo rappresentazionale secondario deve essere possibile. Per esempio, potremmo essere tentati di supporre che il contenuto informativo su un particolare argomento possa sorgere in qualche tratto specializzato e poi, in qualche modo, essere propagato attraverso le regioni corticali, sfruttando la variabilità di queste regioni senza impegnare le semantiche specializzate delle unità ivi residenti. Supponiamo, per esempio, che si verifichi un repentino cambiamento nel quadrante superiore sinistro del mondo visivo di una persona. Come prevedibile, si

potrà notare che l'eccitazione cerebrale sorge prima in quelle parti della corteccia visiva che rappresentano (alla maniera della Camera dei Deputati) varie caratteristiche di eventi nel quadrante superiore sinistro della visione, ma questi punti attivi diventano immediatamente fonti per un'attivazione contagiosa, che implica agenti corticali con altri «elettori». Se questa diffusione dell'eccitazione su varie aree della corteccia non è solo una dispersione o un rumore, se gioca qualche ruolo cruciale nell'elaborazione o revisione di una versione di un frammento narrativo, questi agenti reclutati devono giocare un ruolo piuttosto differente dai ruoli che svolgono quando sono ancorati alla loro fonte.[4]

Non c'è da sorprenderci che non abbiamo, ancora, nessun buon modello di una simile funzionalità multipla (le uniche riflessioni plausibili mi sembrano alcune di quelle di Minsky nella *Società della Mente*). Come abbiamo notato nel capitolo 7, gli ingegneri umani, con la loro imperfetta capacità previsionale, si abituano a progettare sistemi in cui ogni elemento gioca un unico ruolo, isolandolo attentamente da ogni interferenza esterna, allo scopo di minimizzare le conseguenze devastanti degli effetti collaterali imprevisti. Madre Natura, d'altra parte, non si preoccupa affatto degli effetti collaterali imprevisti e quindi può capitalizzare i sorprendenti e positivi effetti collaterali che, ogni morte di papa, si possono verificare. Probabilmente, le oscurità nella scomposizione funzionale della corteccia che finora hanno resistito all'attacco dei neuroscienziati sono dovute alla loro difficoltà costituzionale a concepire ipotesi che assegnano ruoli multipli agli elementi disponibili. Alcuni romantici – il filosofo Owen Flanagan (1991) li chiama i Nuovi Paladini del Mistero – hanno affermato che il cervello incontra un'insormontabile barriera nel comprendere la propria organizzazione (Nagel, 1986, e McGinn, 1990). Non sto avanzando nessuna ipotesi del genere, sto solamente dicendo che si sta dimostrando terribilmente difficile – ma non impossibile – capire come funziona il cervello, in parte perché è stato progettato da un processo che può prosperare grazie a una funzionalità multipla e sovrapposta, qualcosa di sistematicamente difficile da discernere dalla prospettiva dell'ingegneria alla rovescia.

Problemi del genere o passano completamente inosservati o provocano ingiustificate riflessioni a ruota libera. Alcuni possono essere tentati di liquidare su due piedi l'idea di una simile duplicità specialista/generico – non perché possono dimostrare che è errata, ma perché non sanno come inserirla in un modello

e perciò sperano, anche ragionevolmente, di non dover mai averci a che fare. Ma una volta che tale possibilità viene prospettata, è perlomeno in grado di fornire ai teorici qualche indizio utile per la loro ricerca. I neurofisiologi hanno (più o meno) identificato dei meccanismi neuronali, come i recettori NMDA e le sinapsi di von der Malsburg (1985), che potrebbero essere dei candidati plausibili al ruolo di modulatori rapidi della connettività tra cellule. Tali porte potrebbero permettere la rapida formazione di «assemblee» transitorie, che potrebbero sovrapporsi alla rete senza richiedere alcuna alterazione dei legami sinaptici a lungo termine che vengono generalmente ritenuti la colla che tiene unite le assemblee permanenti della memoria a lungo termine. (Per delle nuove riflessioni su questo genere, vedi Flohr, 1990.)

Su scala più grande, i neuroanatomisti stanno compilando la mappa delle connessioni cerebrali, mostrando non solo quali aree sono attive in determinate circostanze, ma iniziando anche a mostrare che tipo di contributo esse offrono. Per varie aree è stato ipotizzato un ruolo fondamentale per la coscienza. Da molto tempo si sa che la *formazione reticolare* nel cervello centrale e il *talamo* sopra di essa giocano un ruolo cruciale nell'attivare la vigilanza cerebrale – all'uscita dal sonno, per esempio, o in risposta a novità o emergenze – e ora che i percorsi sono meglio individuati è possibile formulare e mettere alla prova ipotesi più dettagliate. Crick (1984), per esempio, propone che le ramificazioni che dal talamo si irraggiano verso ogni parte della corteccia lo rendono adatto al ruolo di «riflettore» che risveglia e potenzia le particolari aree specialistiche, reclutandole per gli scopi correnti.[5] Baars (1988) ha elaborato un'idea simile: l'ERTAS o Extended Reticular Thalamic Activating System (Sistema attivante talamico reticolare esteso). Sarebbe abbastanza facile incorporare un'ipotesi del genere nella nostra descrizione appena abbozzata della competizione tra coalizioni di specialisti, a patto di non cadere nell'allettante immagine di un Boss talamico che *capisce* gli avvenimenti attualmente in corso nelle varie parti del cervello con cui è «in comunicazione».

Similmente, si sa che i *lobi frontali* della corteccia, la parte del cervello più cospicuamente sviluppata nell'*Homo sapiens*, siano implicati nel controllo a lungo termine e nella programmazione e sequenzializzazione del comportamento. Un danno in varie zone dei lobi frontali produce tipicamente dei sintomi opposti come la distraibilità e l'incapacità ad abbandonare inveterate abitudini, e l'impulsività o l'incapacità a intrapren-

dere corsi d'azione che richiedono una gratificazione differita. Così potremmo essere tentati di installare il Boss nei lobi frontali, e molti modelli si muovono in questa direzione. Un modello particolarmente sofisticato è il Supervisory Attentional System (sistema di supervisione attenzionale) di Norman e Shallice (1985), che lo collocano nella corteccia prefrontale e a cui danno la responsabilità particolare di risolvere i conflitti quando le burocrazie sussidiarie non possono cooperare. Ancora una volta, trovare una localizzazione anatomica per processi cruciali nel controllo di ciò che avviene poi è una cosa, localizzare il Boss è un'altra; chiunque andasse alla ricerca dello schermo frontale sul quale il Boss segue i progetti da lui controllati sarebbe su una pista falsa (Fuster, 1981; Calvin, 1989a.)

Una volta giurato di abbandonare queste immagini allettanti, però, bisogna trovare altri modi di pensare ai contributi offerti da queste aree; e su questo ci troviamo a corto di idee, malgrado i recenti progressi. Non è che non abbiamo nessuna idea di cosa sia il macchinario; il guaio è che ci manca un modello computazionale di quello che fa il macchinario, e come. Qui stiamo ancora allo stadio della metafora e delle congetture a ruota libera, ma non è uno stadio da evitare; è uno stadio attraverso cui dobbiamo passare nel nostro cammino verso modelli più espliciti.

4. *I poteri della macchina joyceana*

Secondo il nostro schizzo, esiste una competizione tra i vari eventi portatori di contenuto che si verificano contemporaneamente nel cervello, e solo un sottoinsieme scelto di essi «vince»; cioè, riesce a generare effetti duraturi di vario tipo. Alcuni, unendosi ai folletti del linguaggio, contribuiscono alle successive espressioni verbali, sia quelle pronunciate ad alta voce agli altri sia quelle rivolte silenziose a se stessi. Alcuni prestano i loro contenuti ad altre forme di auto-stimolazioni successive, come il «disegnare a se stessi». I rimanenti si estinguono quasi subito, lasciando solo deboli tracce – prove indiziarie indirette – della loro breve esistenza. Qual è il vantaggio, ci si potrebbe chiedere, che alcuni contenuti ottengono nel riuscire ad entrare in questo modo nel «cerchio magico» – e che cosa c'è di tanto magico in *questo* cerchio? La coscienza viene ritenuta qualcosa di straordinariamente speciale. Che cosa c'è di tanto speciale

nell'essere promosso al cerchio successivo in questo ciclo di auto-stimolazione? A che serve? Gli eventi che si verificano in tali meccanismi si rivestono di poteri quasi magici?

Io ho evitato di affermare che una qualche vittoria particolare in questo turbinio competitivo equivalga all'assunzione nella coscienza. In effetti, ho insistito nel dire che non c'è alcun motivo giustificato per tracciare una linea che divida gli eventi che sono definitivamente «nella» coscienza da quelli che stanno per sempre «fuori» o «sotto» la coscienza. (Vedi Allport, 1988, per ulteriori argomenti in favore di questa posizione.) Tuttavia, se la mia teoria della macchina joyceana dovrà gettar luce sulla coscienza, è bene che ci sia qualcosa di notevole su alcune, se non tutte, le attività di questa macchina, perché è innegabile che la coscienza sia, intuitivamente, qualcosa di speciale.

È difficile occuparsi di questi noti interrogativi senza cadere nella trappola di pensare che prima dobbiamo capire *a che cosa* serva la coscienza, per poi chiederci se i meccanismi proposti riescano a realizzare *quella* funzione – qualunque siamo riusciti a stabilire che sia.

Nel suo autorevole libro *Vision* (1982), il neuroscienziato e ricercatore dell'IA David Marr ha proposto tre livelli di analisi che dovremmo cercare di perseguire nella spiegazione di ogni fenomeno mentale. Il livello superiore o più astratto, quello computazionale, è un'analisi «*del problema* [il corsivo è mio] come un compito di elaborazione dell'informazione», mentre il livello mediano, algoritmico, è un'analisi dei processi reali tramite i quali *questo* compito di elaborazione dell'informazione viene svolto. Il livello inferiore, quello fisico, fornisce un'analisi del meccanismo neurale e mostra come esegue l'algoritmo descritto al livello mediano, svolgendo quindi il compito descritto astrattamente al livello computazionale.[6]

I tre livelli di Marr possono essere usati anche per descrivere cose molto più semplici delle menti, ed è possibile farci un'idea delle differenze tra loro vedendo come si applicano a qualcosa così semplice come un abaco. Il suo compito computazionale è di svolgere calcoli aritmetici: arrivare ad un risultato corretto per ogni problema aritmetico dato come input. A questo livello, allora, un abaco ed una calcolatrice tascabile sono simili; sono progettati per svolgere lo stesso «compito di elaborazione dell'informazione». La descrizione algoritmica dell'abaco è quella che impari quando apprendi a manipolarlo – la ricetta per muovere le palline nel corso delle addizioni,

sottrazioni, moltiplicazioni e divisioni. La sua descrizione fisica dipende da quello di cui è fatto: potrebbe essere fatto di palline di legno infilate su fili tesi in un telaio o gettoni da poker allineati sulle giunture tra le mattonelle del pavimento o potrebbe essere realizzato da una matita, una buona gomma e un foglio di carta quadrettato.

Marr raccomanda di costruire modelli dei fenomeni psicologici a tutti e tre questi livelli di analisi e sottolinea particolarmente l'importanza di raggiungere una chiarezza sul livello superiore, computazionale, prima di gettarsi a capofitto sui modelli dei livelli inferiori.[7] Il suo lavoro sulla visione ha brillantemente messo in luce le potenzialità di questa strategia; e da allora altri ricercatori l'hanno utilizzata proficuamente per altri fenomeni. Si potrebbe avere la tentazione di applicare questi tre livelli di analisi alla coscienza come un tutto, e qualcuno ha ceduto alla tentazione. Come abbiamo visto nel capitolo 7, comunque, questa è una rischiosa semplificazione: chiedendoci «Qual è *la* funzione della coscienza?» noi presupponiamo che ci sia un unico «compito di elaborazione dell'informazione» (per quanto complesso) che il meccanismo neurale della coscienza è ben progettato a svolgere – dall'evoluzione, presumibilmente. Ciò ci può portare a trascurare importanti possibilità: alcune caratteristiche della coscienza potrebbero avere molteplici funzioni; alcune funzioni potrebbero essere servite imperfettamente dalle caratteristiche esistenti, a causa delle limitazioni storiche del loro sviluppo; alcune caratteristiche potrebbero non avere affatto delle funzioni – almeno nessuna a *nostro* beneficio. Stando attenti ad evitare queste sviste, allora, passiamo in rassegna i *poteri* (non necessariamente le funzioni) dei meccanismi descritti nel mio schizzo in miniatura.

Prima di tutto, come abbiamo visto nel capitolo 7, ci sono significativi problemi di auto-controllo creati dalla proliferazione di specialisti simultaneamente attivi e uno dei compiti fondamentali svolti dalle attività della macchina joyceana è di giudicare le dispute, smussare i rovesciamenti di regime e prevenire intempestivi *coups d'état* schierando le forze «adatte». Compiti semplici o completamente acquisiti possono essere eseguiti in modo abitudinario senza seri conflitti e senza arruolamento di nuove forze, quindi inconsciamente, ma quando un compito è difficile o spiacevole, esso richiede «concentrazione», cosa che «noi» otteniamo grazie a molte auto-ammonizioni e vari altri trucchi mnemonici, ripetizioni (Margolis, 1989) e altre

auto-manipolazioni (Norman e Shallice, 1985). Spesso ci accorgiamo che ci aiuta parlare ad alta voce, una regressione alle rozze ma efficaci strategie di cui i nostri pensieri privati sono i sofisticati discendenti.

Queste strategie di autocontrollo ci permettono di governare i nostri processi percettivi in maniere che aprono nuove opportunità. Come lo psicologo Jeremy Wolfe (1990) sottolinea, sebbene i nostri sistemi visivi siano spontaneamente finalizzati al rilevamento di certi generi di cose – quelli che «spuntano fuori» quando noi «guardiamo e basta» – ci sono altri generi di cose che possiamo identificare solo se *andiamo deliberatamente alla loro ricerca*, una prassi stabilita da un atto di auto-rappresentazione. Una chiazza rossa in un ammasso di macchie verdi risalterà immediatamente come una chiazza di sangue sulla neve (o meglio, una bacca matura in mezzo al fogliame), ma se i tuoi progetti ti richiedono di trovare una chiazza rossa in mezzo a tante chiazze di altri colori, devi *darti* un compito di ricerca seriale. E se il tuo progetto è quello di trovare il piccolo coriandolo rosso e quadrato in mezzo agli innumerevoli altri di tanti colori e forme differenti (o di rispondere alla domanda «Dov'è Ubaldo?» [Handford, 1987] nell'immagine di un puzzle molto popolare negli Stati Uniti), il compito di ricerca seriale può diventare un progetto metodico che assorbe molta attenzione e richiede un alto grado di auto-controllo.

Queste tecniche con cui rappresentiamo cose a noi stessi ci permettono di autogovernarci o autodirigerci in una maniera che ci allontana molto dalle altre creature. Noi possiamo elaborare delle linee di condotta molto in anticipo, grazie alla nostra capacità di ragionare ipoteticamente e di tessere scenari; possiamo incoraggiarci da soli ad intraprendere progetti spiacevoli o a lungo termine, grazie ad abitudini di auto-rammemorazione e ripetizione dei costi e benefici che ci attendiamo dalle linee di condotta intraprese. E, cosa ancora più importante, questa pratica di ripetizione crea una memoria del percorso fatto per arrivare al punto in cui siamo (quello che gli psicologi chiamano memoria *episodica*), così che possiamo spiegare a noi stessi, quando ci troviamo in un vicolo cieco, quali errori abbiamo fatto (Perlis, 1991). Nel capitolo 7 abbiamo visto come lo sviluppo di queste strategie permise ai nostri antenati di prevedere meglio il futuro, e ciò che ha contribuito a dar loro questo rafforzato potere previsionale fu anche un rafforzato potere di ricordare – essere capaci di guardare più a fondo nelle

proprie recenti operazioni per *vedere* dove si sono fatti errori. «Bene, non dovrò mai più fare *questo*!» è il ritornello di ogni creatura che impara dall'esperienza, ma noi possiamo lanciare i nostri *questo* più indietro e più approfonditamente di ogni altra creatura, grazie alla nostra abitudine ad accumulare passato – o più accuratamente, grazie alla nostra abitudine all'autostimolazione, che ha, tra gli altri effetti, quello di potenziare la capacità di ricordare.

Ma questo caricamento della memoria è solo uno degli effetti preziosi di queste abitudini. Altrettanto importante è l'effetto «diffusione» (Baars, 1988), che crea una specie di forum aperto, che permette a *ogni* cosa che abbiamo appreso di contribuire ad *ogni* problema attuale. Baars sostiene che questa accessibilità mutua dei contenuti fornisce il *contesto* senza il quale gli avvenimenti che si verificano «nella coscienza» non avrebbero – *non potrebbero avere* – senso per il soggetto. I contenuti che compongono il contesto circostante non sono essi stessi sempre coscienti – infatti, in generale essi non sono affatto accessibili, malgrado siano attivati – ma le connessioni tra loro e i contenuti che possono comparire nei resoconti verbali sono ciò che assicura quello che potremmo chiamare il loro significato «consciamente appreso».

Nello stesso spirito, Ray Jackendoff (1987) afferma che i livelli più alti di analisi svolti dal cervello, ossia i più astratti, *non* sono accessibili nell'esperienza, anche se rendono l'esperienza possibile, rendendola significativa. La sua analisi ci fornisce così un utile antidoto contro un'ulteriore incarnazione del Teatro Cartesiano come il «culmine» o «la punta dell'iceberg». (Ecco un buon esempio, tratto dal neuropsicologo Roger Sperry: «In una posizione di comando superiore ai più alti livelli della gerarchia dell'organizzazione cerebrale, le proprietà soggettive [...] esercitano il controllo sulle attività biofisiche e chimiche ai livelli subordinati» [1977, p. 117].)

Non pochi filosofi, specialmente quelli influenzati dalla scuola fenomenologica husserliana (Dreyfus, 1979; Searle, 1983), hanno sottolineato l'importanza di questo «sfondo» di esperienza cosciente, ma l'hanno generalmente descritta come una caratteristica misteriosa o refrattaria, che sfida una spiegazione meccanica, piuttosto che la chiave, come suggeriscono Baars e Jackendoff, per fornire una teoria computazionale di ciò che avviene. Questi filosofi hanno supposto che la coscienza sia la *fonte* di una qualche sorta speciale di «intenzionalità intrin-

seca», ma, come il filosofo Robert van Gulick ha notato, ciò non fa che riportare indietro le cose.

L'esperienza a livello personale della comprensione non è [...] un'illusione. Io, il soggetto personale dell'esperienza, comprendo effettivamente. Posso fare tutte le connessioni necessarie all'interno dell'esperienza, richiamare le rappresentazioni per connetterle immediatamente le une con le altre. Il fatto che la mia abilità sia il risultato del mio essere un sistema organizzato di componenti subpersonali che producono il mio flusso ordinato di coscienza non contraddice la mia abilità. Ciò che è illusorio o sbagliato è solo l'idea che io sia un qualche distinto sé sostanziale che produce queste connessioni in virtù di una forma completamente non comportamentale di comprensione. (van Gulick, 1988, p. 96)

Ogni cosa tra quelle che hai imparato può contribuire a ognuna delle cose con le quali ti stai correntemente confrontando. Questo almeno è l'ideale. Questa caratteristica è chiamata *isotropia* da Fodor (1983), il potere, come direbbe Platone, di far venire gli uccelli appropriati, o almeno di farli cantare, ogni qualvolta si ha bisogno di loro. Sembra magico, ma come ogni prestigiatore sa, la sensazione di trovarsi di fronte ad una magia è rafforzata dal fatto che c'è un pubblico su cui generalmente si può contare per esagerare il fenomeno che deve essere spiegato. Potrebbe sembrare a prima vista che siamo idealmente isotropici, ma non lo siamo. Una riflessione più sobria ci ricorda di tutte le occasioni in cui siamo lenti nel riconoscere la significanza di nuovi dati. Pensa alla classica amplificazione teatrale della situazione di doppio senso in cui ci troviamo quando comprendiamo qualcosa con ritardo o in due tempi (Neisser, 1988). Qualche volta noi addirittura tagliamo il ramo su cui siamo seduti o accendiamo un fiammifero per sbirciare nel deposito del gas.[8]

I prestigiatori sanno che una serie di facili trucchi sono spesso sufficienti per produrre «la magia», e lo sa anche Madre Natura, la grande creatrice di aggeggi magici. I ricercatori dell'Intelligenza Artificiale stanno esplorando lo spazio dei possibili trucchi, ricercando «un fascio di [...] procedure euristiche appropriatamente coordinate e rapidamente impiegate» (Fodor, p. 116) che possa fornire il grado di isotropia che noi pensatori umani dimostriamo. Modelli come ACT* e Soar – e molte altre prospettive indagate nell'IA – sono promettenti ma inconcludenti. Alcuni filosofi, e in particolar modo Dreyfus,

Searle, Fodor e Putnam (1988), sono sicuri che questa idea della Mente come Aggeggio Magico sia sbagliata, e hanno provato a costruire argomentazioni che dimostrano l'impossibilità del compito (Dennett, 1988b, 1991c). Fodor, per esempio, sottolinea che sebbene dei sistemi specialistici possano essere cablati rigidamente, in un sistema a scopi generici che può rispondere con versatilità ad ogni nuova situazione incontrata, «può essere la connettività *instabile* e *istantanea* quella che conta» (p. 118). Egli dispera che chiunque possa presentarsi con una teoria di questa connettività, ma non è solo pessimista: egli dispera *in linea di principio* (un grazioso espediente). Egli ha ragione a supporre che la nostra tendenza all'isotropia sia dovuta al nostro software e non alla nostra cablatura rigida, ma il suo argomento contro l'ipotesi della coscienza come «mucchio di trucchi» presuppone che noi siamo migliori nel «considerare ogni cosa» di quanto non siamo in realtà. Siamo ben fatti, ma non fantastici. Le abitudini di autostimolazione che sviluppiamo ci trasformano in scaltri esploratori delle nostre risorse duramente conquistate; non riusciamo sempre a far cantare l'uccello giusto al momento giusto, ma spesso ci riusciamo e questo fa di noi una ciurma piuttosto buona.

5. *Ma questa è una teoria della coscienza?*

Finora non mi sono sbilanciato molto sulla coscienza; ho attentamente evitato di dire che cosa *sia* la coscienza secondo la mia teoria; non ho dichiarato che qualunque realizzazione di una macchina joyceana sia cosciente e neppure che qualunque stato particolare di tale macchina virtuale sia uno stato cosciente. La ragione della mia reticenza è solo tattica: ho voluto evitare di avviare una rissa sulla coscienza prima di aver avuto la possibilità di far vedere che almeno molti dei presunti poteri della coscienza potrebbero essere spiegati dai poteri della macchina joyceana *indipendentemente* dalla considerazione se una macchina joyceana doni o no la coscienza all'hardware che la ospita.

Potrebbe esistere un essere *in*conscio con uno spazio di lavoro interno e globale in cui numerosissimi folletti si scambiano messaggi, formano coalizioni e tutto il resto? Se sì, allora lo strabiliante potere che l'uomo ha di adattare rapidamente e versatilmente il proprio stato mentale in risposta a quasi tutte le circostanze, per quanto inaspettate, non deve nulla alla co-

scienza in sé, ma solo alla sua architettura computazionale che rende possibile questa collaborazione. Se la coscienza è qualcosa oltre e al di sopra della macchina joyceana, allora io non ho affatto fornito una teoria della coscienza, sebbene siano state risolte altre questioni enigmatiche.

Fino a che non fossero state assemblate tutte le parti del mio abbozzo di teoria, dovevo schivare questi interrogativi, ma ora è giunto il momento di prendere il toro per le corna e confrontarsi con la coscienza in se stessa, l'intero meraviglioso mistero. E quindi io dichiaro ora formalmente che sì, la mia teoria è una teoria della coscienza; ognuno e ogni cosa che ha una macchina virtuale del genere come suo sistema di controllo è cosciente nel senso più pieno del termine ed è cosciente *perché* ha una tale macchina virtuale.[9]

Ora sono pronto ad affrontare le obiezioni. Possiamo cominciare con la questione non risolta di due capoversi fa: qualcosa di inconscio – uno zombi, per esempio – non potrebbe avere una macchina joyceana? Questo interrogativo implica un'obiezione che è così perennemente popolare in momenti come questi che il filosofo Peter Bieri (1990) l'ha battezzata *La ruota tibetana della preghiera*. Non fa altro che ripresentarsi, di volta in volta, senza riguardo a quale teoria sia stata proposta:

Tutto questo va benissimo, potrei essere d'accordo su tutti quei dettagli su come il cervello fa questo e quello, ma io posso immaginare che *tutto ciò* avvenga in un'entità senza che ci sia alcuna vera coscienza!

Una buona risposta a quest'affermazione, ma che raramente si sente, sarebbe: Ah sì, puoi immaginarlo? E come lo sai? Come sai che hai immaginato «tutto ciò» in modo sufficientemente dettagliato e con un'attenzione sufficiente a tutte le implicazioni? Che cosa ti fa pensare che la tua affermazione sia una premessa che conduce a qualche interessante conclusione? Considera quanto saremmo poco impressionati se un *vitalista* attuale ci dicesse:

Tutto questo va benissimo, tutte quelle cose sul DNA e le proteine e simili, ma io posso immaginare di scoprire un'entità che sembri e si comporti proprio come un gatto, fin nei minimi dettagli del sangue nelle sue vene e del DNA nelle sue «cellule», ma che non sia vivo. (Posso davvero? Certo: eccolo lì, che miagola, ma poi Dio mi sussurra nell'orecchio: «Non è vivo! È

solo un DNA-comesichiama meccanico!». Nella mia immaginazione, io credo a Lui.)

Sono certo che nessuno pensi che questo sia un buon argomento a favore del vitalismo. Questo sforzo d'immaginazione non conta. Perché no? Perché è troppo piccino per valere contro la spiegazione della vita offerta dalla biologia contemporanea. La sola cosa che mostra questo «argomento» è che puoi ignorare «tutto ciò» e rimanere attaccato a una convinzione se sei determinato a farlo. La ruota tibetana della preghiera può essere un argomento migliore contro la teoria che ho abbozzato?

Ci troviamo ora nella posizione, grazie all'ampliamento della nostra immaginazione ottenuta nei capitoli precedenti, di spostare l'onere della prova. La ruota tibetana della preghiera (e di essa ci sono varie versioni molto differenti, come vedremo) discende dal famoso argomento di Cartesio (vedi capitolo 2) in cui egli sostiene di essere capace di concepire in modo *chiaro e distinto* che la sua mente è differente dal suo cervello. La forza di un argomento del genere dipende criticamente dallo standard di chiarezza e distinzione con cui giudichiamo la concezione in esame. Alcune persone possono sostenere di poter concepire in modo chiaro e distinto un numero primo massimo o un triangolo che non è una figura rigida. Si sbagliano, o in ogni caso, qualunque cosa stiano facendo quando dicono che stanno concependo queste cose non dovrebbe essere preso come un segno di ciò che è possibile. Ci troviamo ora nella posizione di immaginare «tutto ciò» un po' più dettagliatamente. Puoi *realmente* immaginare uno zombi? L'unico senso in cui è «ovvio» che puoi farlo è un senso che non mette in discussione la mia teoria, e un senso più forte, meno ovvio, ha bisogno di una dimostrazione.

I filosofi non hanno, di regola, chiesto tanto. I più influenti esperimenti di pensiero nella recente filosofia della mente hanno tutti invitato il pubblico a immaginare un certo stato di cose specificamente stabilito o escogitato, per poi – senza controllare adeguatamente se questo sforzo d'immaginazione è stato effettivamente realizzato – invitare il pubblico a «rimarcare» le varie conseguenze della fantasia. Queste «pompe d'intuizione», come io le chiamo, sono spesso dei congegni terribilmente astuti. Meritano la loro fama, non fosse altro che per il loro fascino.

Nella terza parte le affronteremo tutte, sviluppando la nostra teoria della coscienza man mano che andremo avanti.

Dalla nostra nuova prospettiva, saremo in grado di vedere l'abile colpo di mano che confonde il pubblico – e l'illusionista – e nel contempo affineremo i nostri poteri d'immaginazione. Tra i famosi argomenti che incontreremo ci sono non solo «la presunta possibilità degli zombi» ma anche «lo spettro invertito», «quello che Mary la scienziata del colore non sa sul colore», «la stanza cinese» e «che cosa si prova ad essere un pipistrello».

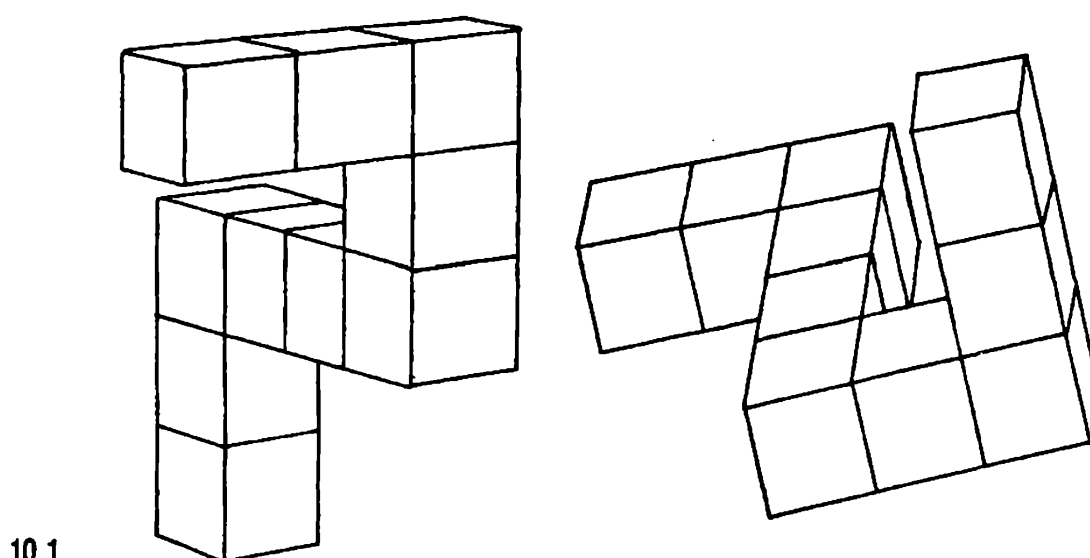
PARTE TERZA

I PROBLEMI FILOSOFICI DELLA COSCIENZA

10

MOSTRARE E DIRE*

1. *La rotazione di immagini con l'occhio della mente*

Prima di esaminare gli esperimenti di pensiero filosofici, dobbiamo affrontare alcuni esperimenti reali che sembrerebbero riabilitare il Teatro Cartesiano. Alcune delle ricerche più stimolanti e ingegnose degli ultimi venti anni nelle scienze cognitive sono state quelle sulla capacità umana di manipolare immagini mentali, avviate dal classico studio dello psicologo Roger Shepard (Shepard e Metzler, 1971) sulla *velocità di rotazione mentale* di figure come queste:

10.1

* Come ogni lettore americano sa, *Show and Tell*, che potrebbe essere tradotto anche con «mostra e descrivi» o «mostra e di'», allude alla pratica molto diffusa nelle scuole elementari americane di invitare gli alunni a portare, a turno, un oggetto strano, curioso, interessante da mostrare ai compagni e da descrivere a parole, unendo così il modo ostensivo e quello descrittivo di riferirsi a qualcosa. Non si dimentichi inoltre l'altro significato di *show*: spettacolo – che di solito si svolge in un teatro (cartesiano?). [*N.d.T.*]

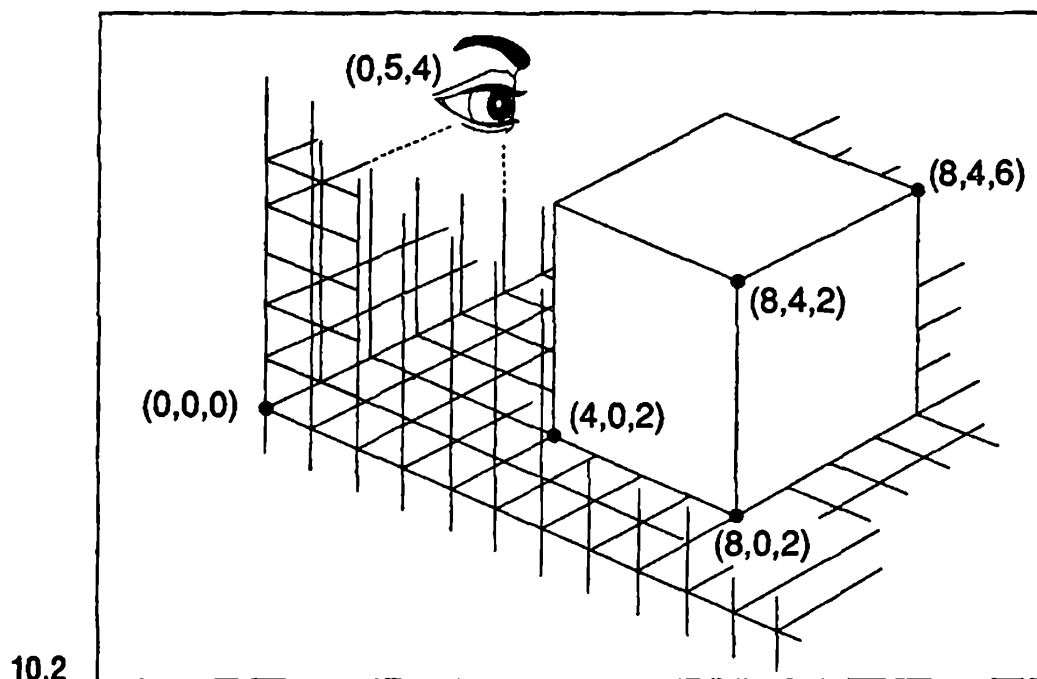
Nell'esperimento originale, ai soggetti veniva mostrata una coppia di disegni del genere e veniva chiesto se tale coppia rappresentasse un'unica forma vista da posizioni differenti. In questo caso, come puoi constatare rapidamente, la risposta è sì. Come lo hai stabilito? Una risposta tipica è: «Facendo ruotare una delle immagini con l'occhio della mente e sovrapponendola all'altra». Shepard variò le distanze angolari di rotazione tra le coppie di figure – alcune coppie avevano bisogno di una rotazione di pochi gradi per essere allineate, altre richiedevano rotazioni molto più ampie – e misurò il tempo che i soggetti impiegavano, mediamente, nel rispondere alle differenti immagini mostrate. Supponendo che nel cervello si verifichi qualcosa di simile ad un processo di effettiva rotazione di immagini, si dovrebbe impiegare il doppio del tempo a ruotare la stessa immagine di 90 gradi invece che di 45 gradi (se trascuriamo l'accelerazione e la decelerazione, mantenendo la velocità di rotazione costante).[1] I dati di Shepard confermano molto bene questa ipotesi, anche in condizioni molto differenziate. Centinaia di esperimenti successivi, di Shepard ed altri, hanno indagato dettagliatamente il comportamento del meccanismo cerebrale per la manipolazione delle immagini e – per esprimermi nel modo più cauto possibile su un argomento ancora controverso – sembra davvero che nel cervello esista quello che lo psicologo Stephen Kosslyn (1980) chiama un «buffer visivo» o «memoria di transito visiva», che esegue trasformazioni tramite processi che sono fortemente «immaginistici» – o, per usare il termine di Kosslyn, *quasi-pittorici*.

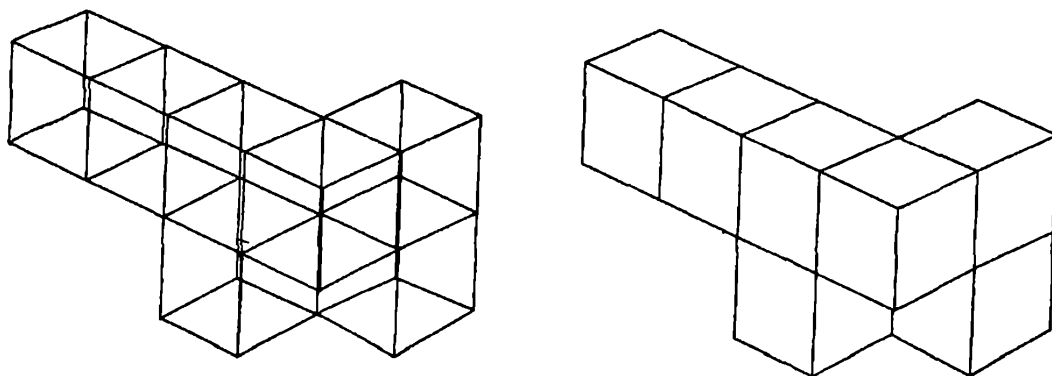
Che significa questo? Gli psicologi cognitivi hanno forse scoperto che il Teatro Cartesiano esiste dopo tutto? Secondo Kosslyn, questi esperimenti mostrano che le immagini vengono assemblate per l'esposizione interna in una maniera che assomiglia molto a quella in cui le immagini su un CRT (un tubo a raggi catodici come uno schermo televisivo o un monitor di calcolatore) possono essere create a partire dai file nella memoria di un calcolatore. Una volta che si trovano sullo schermo interno, esse possono essere esaminate, ruotate e altrimenti manipolate dai soggetti a cui sono stati dati particolari compiti da svolgere. Kosslyn sottolinea, comunque, che il suo modello CRT è una metafora. Questo ci dovrebbe far ricordare di un'altra metafora, quella connessa alla capacità di «manipolazione di immagini» di Shakey. Certamente Shakey non aveva nessun Teatro Cartesiano nel suo cervello di silicio. Per ottenere un quadro un po' più chiaro di quello che deve effettivamente

avvenire in un cervello umano, possiamo iniziare con un modello *non* metaforico, troppo forte per essere vero, e poi «sottrarre» le proprietà indesiderate dal modello ad una ad una. In altre parole, accettiamo la metafora del CRT di Kosslyn e poi correggiamola gradualmente con delle limitazioni.

Come prima cosa si consideri un sistema che manipola realmente immagini reali, come i sistemi di grafica al calcolatore che oggigiorno proliferano in centinaia di adattamenti: animazioni al calcolatore per film e televisione, sistemi che rappresentano oggetti tridimensionali da vari punti di vista per architetti e arredatori, videogiochi e molti altri. Gli ingegneri chiamano le loro versioni sistemi CAD, per *computer-aided design* (progettazione assistita dal calcolatore). I sistemi CAD stanno rivoluzionando la progettazione non solo perché rendono enormemente più semplice il disegnare, come la videoscrittura rende enormemente più semplice lo scrivere, ma anche perché grazie ad essi è possibile risolvere facilmente problemi e rispondere a interrogativi che sarebbero altrimenti molto ostici. Messo di fronte al problema di Shepard della figura 10.1, un ingegnere potrebbe rispondere con l'aiuto di un sistema CAD, mettendo entrambe le immagini sullo schermo CRT, ruotando letteralmente una di esse e provando poi a sovrapporle. Ma attenzione, alcuni dettagli del processo sono importanti.

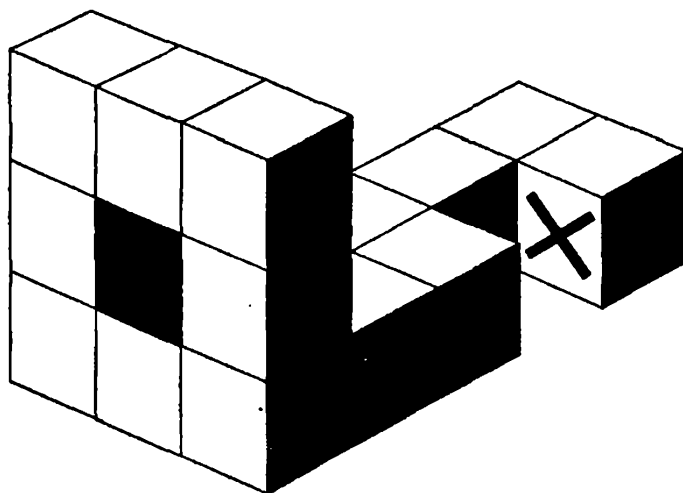
Ogni oggetto disegnato entrerebbe nella memoria del calcolatore come un oggetto tridimensionale *virtuale*, grazie a un'analisi che lo ridurrebbe a una descrizione di piani e bordi definiti dalle loro coordinate *xyz*, essendo ogni punto nello spazio virtuale un «tripla ordinata» di numeri immagazzinati nella memoria. Anche il punto di vista dell'osservatore implicito comparirebbe come un punto nello stesso spazio virtuale, definito dalle sue tre coordinate *xyz*. Ecco il disegno di un cubo e di un punto di vista, ma è importante ricordare che l'unica cosa che il calcolatore deve immagazzinare è una tripletta per ogni punto cruciale, raggruppate in insiemi più grandi (per es., uno per ogni faccia del cubo), assieme alle informazioni codificate per le varie proprietà di ogni faccia (il suo colore, l'eventuale opacità o trasparenza, la sua *texture*, e così via). La rotazione e lo spostamento di un oggetto in questo spazio virtuale possono essere facilmente calcolati, semplicemente modificando di una quantità costante tutte le coordinate *x*, *y* e *z* dell'oggetto – pura aritmetica. Infine, è una questione di semplice geometria calcolare le linee prospettiche che determinano quali lati dell'oggetto sarebbero visibili dal punto di vista virtuale e come apparireb-

10.2

bero esattamente. I calcoli sono semplici ma laboriosi o «ad alta intensità di calcolo», specialmente se devono essere calcolate anche le linee curve, l'ombreggiatura, la luce riflessa e la *texture.*

Sui sistemi avanzati è possibile calcolare le differenti schermate con una rapidità sufficiente a creare un movimento apparente, ma a patto di mantenere le rappresentazioni piuttosto schematiche. La «rimozione delle linee nascoste», il processo computazionale che rende opaca nei punti giusti l'immagine finale e impedisce a un cubo di Shepard di sembrare come un cubo di Necker trasparente, è di per sé un processo che consuma un tempo non indifferente, e mette dei limiti a ciò che può essere prodotto «in tempo reale». Nelle meravigliose elaborazioni grafiche di immagini che vediamo ogni giorno alla televisione, il processo di calcolo che genera le immagini è molto più lento, anche su supercomputer, e quindi i singoli fotogrammi

10.3 Prima e dopo la rimozione delle linee nascoste.

devono essere immagazzinati per essere successivamente mostrati a una velocità maggiore che soddisfi i requisiti posti dal sistema visivo umano al rilevamento del movimento.[2]

Questi manipolatori di oggetti virtuali a tre dimensioni sono degli strumenti o giochi nuovi e magnifici, sono davvero qualcosa di nuovo sotto il sole e non copie elettroniche di qualcosa che avevamo già nella nostra testa. È abbastanza certo che nel nostro cervello non si verifica nessun processo analogo a questi triliardi di calcoli geometrici e aritmetici quando giochiamo con le nostre immagini mentali, e niente altro *potrebbe* rendere le sequenze animate così riccamente dettagliate che tali calcoli producono – per le ragioni che abbiamo visto nel capitolo 1.

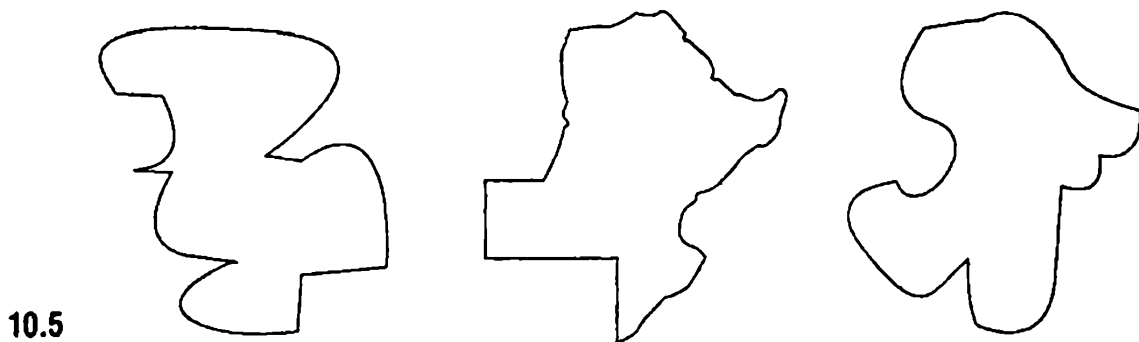
Possiamo convincerci che questo limite del nostro cervello è reale considerando un problema leggermente differente, ma sempre nello stile di Shepard, che sarebbe facilmente risolvibile con l'aiuto di un sistema CAD: la X «rossa» su una faccia di questo oggetto sarebbe visibile per qualcuno che guarda attraverso il buco quadrato nel muro frontale?

10.4

Il nostro oggetto «shepardiano» con la X è un oggetto semplice e schematico, e poiché la nostra domanda non riguarda la *texture*, l'illuminazione e altre simili finezze, sarebbe abbastanza facile per un ingegnere produrre una rotazione animata di questo oggetto su un CRT. Egli potrebbe poi ruotare l'immagine in ogni direzione, spostando il punto di vista avanti e indietro – alla ricerca di un barlume rosso attraverso il buco. Se vede il rosso, la risposta è sì, altrimenti no.

Ora, puoi fare la stessa cosa con l'occhio della tua mente? Puoi semplicemente ruotare l'oggetto mostrato e sbirciare attraverso il buco? Se puoi farlo, riesci a fare qualcosa che io non so fare e che neanche tutte le persone a cui l'ho domandato

riescono a fare con qualche sicurezza. Anche quelli che hanno una risposta sono abbastanza sicuri di poterla dare non perché hanno *solo* ruotato e guardato. (Dicono spesso che prima hanno provato a ruotare e a guardare, ma hanno scoperto che non funziona; possono «ruotare» l'oggetto ma poi esso «va in pezzi» quando provano a guardare attraverso il buco. Poi dicono di aver «tracciato» delle linee prospettiche attraverso il buco sull'immagine *non* ruotata, per vedere se riuscivano a stabilire il punto in cui colpivano il piano retrostante.) Poiché il nostro oggetto shepardiano non è più complesso degli oggetti che vengono ruotati con apparente successo in molti esperimenti, sorge un problema: che tipo di processo può eseguire così facilmente alcune trasformazioni (e poi ricavare informazioni dal risultato) e fallire così malamente in altre operazioni che non sembrano più difficili? (Se queste operazioni non ci *sembrano* più difficili, forse le stiamo considerando da un punto di vista sbagliato, poiché i nostri fallimenti dimostrano che *sono* più difficili.)

Un esperimento degli psicologi Daniel Reisberg e Deborah Chambers (di prossima pubblicazione) solleva lo stesso problema. A soggetti che si ritenevano bravi nel visualizzare immagini mentali furono mostrate delle forme «senza senso» e poi fu chiesto di ruotarle di 90 o 180 gradi con l'occhio della mente e di riferire cosa «vedevano». I soggetti stessi si stupivano di non riuscire a riconoscere, con l'occhio della mente, quello che puoi facilmente riconoscere ruotando il libro in senso orario di 90 gradi per osservare queste figure:

10.5

Le domande alle quali gli ingegneri rispondono usando i sistemi CAD non sono normalmente così semplici come: «Attraverso il buco si può vedere la X rossa?». Di solito sono interessati a proprietà spaziali più complicate degli oggetti che stanno progettando, come: «Questo braccio snodabile in tre parti del robot potrà raggiungere la sua schiena e girare la manopola lì posta senza urtare l'alloggiamento delle batterie?»,

o anche proprietà estetiche, come: «La scalinata nell'ingresso di questo albergo come apparirà a qualcuno che passa per strada e guarda attraverso la vetrata principale?». Quando proviamo a visualizzare queste scene senza un aiuto otteniamo solo dei risultati molto approssimativi e poco affidabili, quindi un sistema CAD può essere considerato come una specie di protesi per l'immaginazione (Dennett, 1982d, 1990b); amplifica enormemente i poteri immaginativi di un essere umano, ma solo a patto che l'utente abbia una vista normale – per guardare lo schermo CRT.

Proviamo ora a immaginare un congegno protesico più ambizioso: un sistema CAD per ingegneri ciechi! E, per non complicare le cose, supponiamo che gli interrogativi che questi ingegneri ciechi vogliono risolvere siano relativamente semplici, del tipo geometrico – e non riguardino valutazioni estetiche. L'output ovviamente dovrà essere in qualche forma non visiva; la forma più facile per l'utente sarà una risposta in linguaggio ordinario (in Braille o in voce sintetizzata) a domande in linguaggio ordinario. Così supporremo che quando si trova di fronte a una domanda come quella che abbiamo appena considerata, l'ingegnere cieco non debba far altro che passare la domanda al sistema CAD (in termini che esso possa «capire», ovviamente) e aspettare che gli fornisca una risposta.

Il nostro sistema CADBLIND Mark I non sarà molto elegante, ma è assai semplice. Consiste di un normale sistema CAD, completo di uno schermo CRT, di fronte al quale sta appollaiato un *Vorsetzer*, un sistema elettronico per la visione, completo di telecamera puntata sul CRT e di dita meccaniche per girare le manopole del sistema CAD.[3] A differenza di Shakey, il cui CRT era solo a beneficio degli osservatori esterni, questo sistema «guarda» davvero l'immagine, un'immagine reale di punti fosforescenti, che irraggiano luce reale di diverse frequenze sui trasduttori sensibili alla luce sul fondo della telecamera. Quando gli si pone il problema shepardiano della X rossa, il CADBLIND Mark I produce un'immagine che contiene una X rossa reale, visibile a tutti, compresa la telecamera del *Vorsetzer*.

Supponiamo, senza ulteriori complicazioni, che il *Vorsetzer* abbia risolto parecchi problemi connessi alla visione al calcolatore e sia quindi in grado di estrarre le informazioni volute dalla rappresentazione luminescente sullo schermo CRT. (No, non intendo sostenere che il *Vorsetzer* sia conscio – voglio solo supporre che sia abbastanza bravo da poter rispondere alle

domande che gli ingegneri ciechi gli pongono.) Il CADBLIND Mark I produce e manipola immagini reali, e le usa per rispondere, per conto degli ingegneri ciechi, a tutte quelle domande a cui un ingegnere vedente potrebbe rispondere usando un normale sistema CAD. Se questo è il sistema Mark I, allora sarà un gioco da ragazzi progettare il sistema Mark II: basta buttare via il CRT e la telecamera che lo inquadra, e sostituirli con un semplice cavo! Attraverso questo cavo il sistema CAD invia al *Vorsetzer* la *mappa di bit*, la matrice di zero e uno che definisce l'immagine sul CRT. Nel *Vorsetzer* del Mark I, questa mappa di bit era laboriosamente ricostruita a partire dai dati in uscita dai trasduttori ottici nella telecamera.

Ci sono scarsi risparmi *computazionali* nel Mark II – solo l'eliminazione di hardware non necessario. Tutti gli elaborati calcoli sulle linee prospettiche, la rimozione delle linee nascoste, la resa della tessitura, le ombre e la luce riflessa, che occupavano uno spazio così grande dei calcoli del Mark I, sono ancora presenti nel processo. Si supponga che il *Vorsetzer* del Mark II sia chiamato a dare un giudizio di profondità comparando i gradienti di *texture* o interpretando le ombre; per farlo, dovrà *analizzare* la configurazione di bit nelle porzioni rilevanti della mappa di bit per giungere alle discriminazioni delle *texture* e delle ombre.

Ciò significa che il Mark II è ancora una macchina ridicolmente inefficiente, giacché se l'informazione che una porzione particolare della mappa di bit rappresenta un'ombra è già «conosciuta» dal sistema CAD (se fa parte della descrizione codificata dell'oggetto con cui il sistema CAD genera le sue immagini) e se questo fatto fa parte di ciò che il *Vorsetzer* deve determinare per arrivare al suo giudizio di profondità, perché il sistema CAD non lo *dice* semplicemente al *Vorsetzer*? Perché preoccuparsi di *rendere* le ombre a beneficio dell'analizzatore di configurazioni del *Vorsetzer*, quando il compito di rendere la configurazione e analizzarla si eliminano a vicenda?

Così il nostro CADBLIND Mark III sarà dispensato dagli ingombranti compiti computazionali di rendere l'immagine grazie a un passaggio diretto delle «conoscenze» sull'oggetto rappresentato al sottosistema *Vorsetzer*, utilizzando il formato di semplici codici per le proprietà e attaccando «etichette» a vari «luoghi» della matrice della mappa di bit, che è così trasformata da una pura immagine in qualcosa di

simile a un diagramma. *Alcune* proprietà spaziali sono rappresentate direttamente – *mostrate* – nello spazio (virtuale) della mappa di bit, ma altre sono solo *dette* tramite etichette.[4]

Questo dovrebbe farci ricordare l'affermazione fatta nel capitolo 5: il cervello deve fare le sue discriminazioni solo una volta; una caratteristica identificata non deve essere riesposta a beneficio di un valutatore capo nel Teatro Cartesiano.

Ma ora possiamo vedere un aspetto ingegneristico differente: la «eliminazione a vicenda» funziona solo se i sistemi che devono comunicare possono «parlare la stessa lingua». Che succede se il formato in cui il sistema CAD «conosce» già le informazioni rilevanti – per es., l'informazione che qualcosa è un'ombra – non è un formato «utilizzabile» dal *Vorsetzer*?[5] Allora, affinché ci sia comunicazione sarà forse necessario «arretrare per andare avanti». Sarà forse necessario che i sistemi realizzino interazioni dispersive in senso informazionale – si potrebbe dire di ampio respiro – affinché possano semplicemente interagire. Si pensi a quando abbozziamo una mappa per dare delle indicazioni stradali a uno straniero, quando tutto quello che egli dovrebbe sapere – se solo sapessimo dirlo nella sua lingua – è «gira a sinistra al prossimo semaforo». Affrontare tutte le lungaggini per produrre qualcosa di simile ad un'immagine è spesso praticamente necessario, anche quando non è necessario «in linea di principio».

Poiché i sistemi nei nostri cervelli sono il prodotto di vari processi sovrapposti di riaggiustamento opportunistico, quello di lungo periodo della selezione naturale e quello di breve periodo della riprogettazione individuale tramite l'autostimolazione, non dovremmo stupirci di trovare tali inefficienze. Inoltre, ci sono altre ragioni per rendere le informazioni in formati simili alle immagini (oltre al semplice piacere di farlo) e, quando ci capita di farlo inavvertitamente, ci accorgiamo subito che esse giustificano ampiamente tutto il lavorio necessario. Come abbiamo già notato, nelle riflessioni del capitolo 7 sul «disegnare a se stessi», tali trasformazioni di formato sono spesso delle maniere molto efficaci di estrarre delle informazioni che sarebbero altrimenti del tutto inestricabili dai dati. I disegni equivalgono effettivamente a una ri-*presentazione* dell'informazione – non a un occhio interno, ma a un meccanismo interno di riconoscimento di configurazioni che può accettare input anche da un occhio normale («esterno»?). Ecco perché le tecniche di elaborazione grafica al calcolatore sono così preziose per la scienza, per esempio; permettono a grandi insiemi ordi-

nati di dati di essere presentati in un formato utilizzabile dalle strepitose capacità di riconoscimento di configurazioni della visione umana. Ci costruiamo schemi e mappe e ogni sorta di grafici codificati con colori per fare in modo che gli aspetti costanti e salienti dei dati «emergano da soli», grazie ai nostri sistemi visivi. I disegni non ci aiutano solamente a vedere configurazioni che sarebbero altrimenti impercettibili; ci aiutano a *tener conto* di ciò che è rilevante e a *ricordarci* di fare le domande giuste al momento giusto. Il ricercatore svedese di IA Lars-Erik Janlert (1985) ha sostenuto che una tale generazione e scrutinio di immagini in un calcolatore può essere usata anche per aiutare a risolvere problemi altrimenti intrattabili di ciò che potremmo chiamare la gestione dell'inferenza in sistemi che sono «in linea di principio» dei motori puramente deduttivi. (Per un altro punto di vista sullo stesso processo, vedi Larkin e Simon, 1987.)

Questa tattica è sicuramente ben conosciuta da molti acuti pensatori ed è stata descritta meravigliosamente da uno dei più acuti di tutti i tempi, il fisico Richard Feynman, nel suo *Sta scherzando, Mr Feynman!* (1985). Nel capitolo giustamente intitolato: «Una differente scatola di attrezzi», egli racconta come stupiva i suoi compagni nei corsi di specializzazione a Princeton «intuendo» la verità o la falsità di arcani teoremi di topologia, teoremi che egli era assolutamente incapace di derivare formalmente o perfino comprendere pienamente:

Applicavo uno schema che utilizzo ancora oggi quando qualcuno spiega una cosa che voglio capire: continuo a fare esempi. Facciamo il caso di matematici che intuiscono un teorema fantastico. Mentre mi dicono le condizioni del teorema, mi costruisco un modello che vi si adatta. Abbiamo un insieme (una boccia, traduco io) [...] disgiunto (due bocce). Poi nella mia testa la boccia prende colore, si copre di peli man mano che aggiungono altre condizioni. Quando poi arriva il teorema, una qualche stupidata che non corrisponde per niente alla boccia verde e pelosa che ho in mente, rispondo «falso».

Se invece è corretto, loro si esaltano, io li lascio andare avanti per un po', poi tiro fuori il mio controesempio.

«Oh, ci siamo scordati di dirti che si tratta di un omomorfismo di Hausdorff di classe 2...»

«Allora, è proprio elementare», dico io, «elementare». A questo punto ho capito come va a finire, anche se non so niente degli omomorfismi di Hausdorff.

Tattiche del genere «sono naturali» in un certo senso, ma

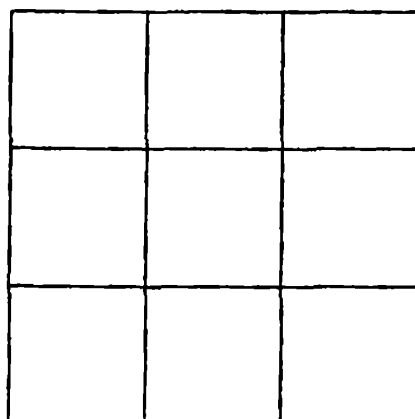
devono essere apprese o inventate, e alcune persone ci riescono meglio di altre. Quelle con abilità più sviluppate hanno nel cervello delle macchine virtuali differenti, con poteri notevolmente differenti da quelli delle persone che «visualizzano» raramente e con difficoltà. E le differenze emergono facilmente nei loro mondi eterofenomenologici individuali.

Allora ci sono buone ragioni per credere, come Kosslyn e altri hanno sostenuto, che gli esseri umani mettano in funzione i loro sistemi visivi non solo per presentare a se stessi le immagini reali, *esterne* (come sul CRT dei sistemi CAD), ma anche immagini virtuali interne o rappresentazioni di dati in forma di disegni che sono provocate endogenamente e in modo personale, e costituiscono un adeguato materiale grezzo per gli stadi ulteriori del meccanismo dell'elaborazione visiva.

Ma esattamente in quali soluzioni ingegneristiche a quali problemi di comunicazione e manipolazione interne di informazione il cervello umano si è imbattuto, e quali sono i loro punti di forza e di debolezza?[6] Queste sono le domande empiriche a cui la ricerca nella scienza cognitiva sull'immaginazione cerca di dare una risposta e noi dovremmo essere molto cauti nel cercare di avanzare delle risposte *a priori*.[7] Immagino che nel nostro cervello avremmo anche potuto trovare un sistema di manipolazione d'immagini come il Mark I, completo di punti fosforescenti e occhio interno. (Per quanto ne so io, non è *impossibile* che le creature di un lontano pianeta siano dotate di simili marchingegni.) E servono esperimenti come quelli di Reisberg e Chambers per mostrare che le scorciatoie imboccate dal nostro cervello precludono molto probabilmente la scoperta di un sistema Mark II, con un formato a mappa di bit che non si avvantaggia mai delle scorciatoie. (Se avessimo un tale sistema, risolvere il quesito della X rossa nella nostra testa sarebbe facile, e lo sarebbe anche ruotare il Texas.)

La fenomenologia ci fornisce dei suggerimenti, che indicano però due direzioni. La «schematicità» delle immagini mentali, che è «intuitivamente ovvia» nella fenomenologia della maggior parte dei soggetti, indica che il cervello usa delle scorciatoie, casi in cui il cervello «dice senza mostrare». Questo è vero sia della *visualizzazione* sia della *percezione*. Abbiamo già notato nel capitolo 3 come è difficile disegnare una rosa che sta proprio di fronte ai nostri occhi o perfino copiare un disegno, e ciò perché le proprietà puramente spaziali che devono essere identificate o discriminate per disegnare bene, normalmente sono state lasciate indietro nel corso dell'elaborazione percettiva, riassunte

in resoconti, non rese per un ulteriore scrutinio. D'altra parte, l'utilità delle immagini mentali nell'aiutarci a «vedere la configurazione» o a «ricordarci» dei dettagli che potremmo altrimenti dimenticare, indica lo sfruttamento del meccanismo visivo di riconoscimento della configurazione, che può verificarsi solo se una parte del cervello si è precedentemente assunta il compito di preparare versioni delle informazioni nel formato speciale utilizzato da quei sistemi visivi. Come già vedemmo nel capitolo 1, i requisiti necessari alla manipolazione delle informazioni di queste rappresentazioni sono formidabili e non dovrebbe sorprenderci se riusciamo così malamente a mantenere stabili nelle nostre menti anche diagrammi fortemente schematici.

Ecco un semplice test per ricordarti quanto sono limitate le nostre capacità: *con l'occhio della mente*, riempi il seguente cruciverba da tre caselle per tre, inserendo verticalmente nelle colonne queste tre parole, partendo da sinistra: GAS ALT PIO

10.6

Puoi *leggere* agevolmente le parole orizzontali? In un effettivo diagramma sulla carta, queste parole «emergerebbero da sole» – ti sarebbe impossibile *non* vederle. A questo, dopotutto, servono i diagrammi: per presentare i dati in un formato che rende inevitabile o facile l'analisi dei dati. Una matrice tre per tre di caratteri alfabetici non è una struttura di dati molto complicata, eppure il nostro cervello non riesce a mantenerla sufficientemente stabile da permettere al sistema visivo di fare il suo lavoro di «percezione» immediata dei dati. (Se vuoi provare di nuovo, ecco altri due gruppi di parole per le colonne: EVA GIN ORO e BAR IVI TEO.)

Esiste comunque un'ampia gamma di differenze individuali nelle tattiche impiegate dalle diverse persone che visualizzano e alcuni possono essere in grado di trovare – o sviluppare – strategie immaginative che permettono loro di «leggere» questi

diagrammi. Gli individui con prodigiose capacità di calcolo possono escogitare metodi per moltiplicare mentalmente tra loro numeri di dieci cifre, e quindi non dovrebbe sorprenderci che alcune persone possano sviluppare prodigiosi talenti per la «lettura di cruciverba» con l'occhio della mente. Queste dimostrazioni informali ci danno degli indizi, ma sono necessari esperimenti che possano definire più incisivamente i tipi di meccanismi e processi che le persone usano in questi atti di auto-manipolazione. Dalle prove disponibili finora sembrerebbe che noi usiamo una strategia mista, sfruttando alcuni vantaggi dell'analisi visiva degli insiemi ordinati, ma utilizzando anche la scorciatoia delle etichette, *dicendo* senza *mostrare*.

Si noti, comunque, che anche per il sistema CADBLIND Mark II, che è ultra-pittorico poiché incorpora la mappa di bit che rende i colori, le ombre e la *texture* pixel per pixel, c'è un senso tuttavia – ed è un senso metafisicamente importante come vedremo nei prossimi due capitoli – in cui tutto è «dire» e niente è «mostrare». Si consideri la X rossa nella nostra figura shepardiana (figura 10.4). Nel Mark I è resa da un rosso reale – il CRT emette luce, che deve essere trasdotta nella telecamera da qualcosa di analogo ai coni nei nostri occhi che rispondono alle differenze di frequenza. Quando il *Vorsetzer* ruota avanti e indietro l'immagine alla ricerca di un barlume rosso attraverso il buco, aspetta che i suoi folletti rilevatori del rosso gli facciano un urlo. Nel Mark II, quell'hardware non c'è più, e la mappa di bit rappresenta il colore di ogni pixel con un numero. Forse la tonalità rossa è il colore numero 37. Il *Vorsetzer* nel Mark II, quando ruota l'immagine della mappa di bit avanti e indietro, è alla ricerca di un barlume di 37; o in altre parole, si chiede se qualche folletto analizzatore di pixel voglia dirgli: «Ecco il colore numero 37!». Tutto il rosso è svanito: ora ci sono solo numeri. E in definitiva, tutto il lavoro in un sistema CADBLIND deve essere fatto tramite operazioni aritmetiche su stringhe di bit, proprio come abbiamo visto nel livello più basso di Shakey nel capitolo 4. E indipendentemente da quanto possano essere quasi-pittorici o immaginistici i processi che terminano nella risposta che il *Vorsetzer* dà alle domande, essi non sono generati in un posto interno dove le proprietà perdute (le proprietà di cui «si parla» solamente nella mappa di bit) vengano *ripristinate* per essere *apprezzate* da un giudice che elabora le risposte.

Le persone non sono sistemi CADBLIND. Il fatto che un sistema CADBLIND possa manipolare e ispezionare le sue

«immagini mentali» senza il beneficio di un Teatro Cartesiano non prova di per sé che nel cervello umano non esista il Teatro Cartesiano, ma dimostra che non dobbiamo postulare un Teatro Cartesiano per spiegare la capacità umana di risolvere i problemi «con l'occhio della mente». Esistono effettivamente dei processi mentali che sono fortemente *analoghi* all'osservazione, ma, quando riduciamo all'essenziale la metafora CRT di Kosslyn, eliminiamo proprio quelle caratteristiche che chiamerebbero in causa un Teatro Cartesiano. Non è necessario che ci sia un tempo e un luogo in cui «tutto confluisce» a beneficio di un unico discriminatore unitario; la discriminazione può essere realizzata in un modo distribuito, asincrono e a molti livelli.

2. *Parole, immagini e pensieri*

> L'aspetto veramente «creativo» del linguaggio risiede non nella sua «infinita capacità generativa», ma in cicli di produzione e comprensione mediati da una mente capace di riflettere sui molteplici significati attribuibili a un'espressione, significati che non devono necessariamente essere presenti nel pensiero che ha dato origine all'espressione ma che diventano disponibili tramite l'auto-comprensione (o l'interpretazione profonda dell'espressione di un altro) e possono condurre a un nuovo pensiero da esprimere e reinterpretare, e così via indefinitamente.
>
> H. STEPHEN STRAIGHT (1976), p. 540

All'economista britannico John Maynard Keynes una volta fu chiesto se pensava in immagini o in parole. «Penso in pensieri», fu la sua risposta. Aveva ragione a resistere alla tentazione di credere che «le cose tramite cui pensiamo» siano o parole o immagini, giacché, come abbiamo visto, le «immagini mentali» non sono *proprio* come immagini o fotografie nella testa e pensare «verbalmente» non è *proprio* come parlare a se stessi. Ma dire che si pensa in pensieri non è davvero un passo avanti; non si fa altro che posporre la questione, giacché un pensiero è proprio *quel qualcosa che si verifica quando pensiamo* – un argomento su cui non c'è un accordo comune.

Ora che abbiamo esaminato le linee generali del tipo di macchinario che da dietro le quinte è responsabile dei dettagli dei nostri mondi eterofenomenologici, possiamo *iniziare* a spiegare la fenomenologia del pensare, prendendo in esame non

solo i limiti e le condizioni della fenomenologia «visiva» e «verbale», ma anche altre eventuali varietà che ci permettono di sfuggire a questa dicotomia.

Uno dei miei esercizi favoriti di eterofenomenologia letteraria si basa sul romanzo di Vladimir Nabokov *The Defense* (1930), incentrato sul Granmaestro Luzhin, un genio della scacchiera, che ha un crollo nervoso nel bel mezzo della partita più importante della sua vita. Possiamo seguire tre stadi nello sviluppo della sua coscienza: la sua mente infantile (prima della scoperta degli scacchi a dieci anni), la sua mente satura del gioco degli scacchi (fino al suo crollo nervoso) e i tristi residui dei due stadi precedenti dopo il crollo, quando, imprigionato dalla moglie in un mondo senza scacchi – senza discussioni sugli scacchi, partite di scacchi, libri di scacchi – la sua mente regredisce ad una specie di viziata paranoia infantile, illuminata da fugaci momenti scacchistici rubati – rapidi attacchi furtivi sugli schemi e i problemi scacchistici nei giornali – ma che prende infine la forma di un'ossessione scacchistica che culmina nel suo «auto-scacco suicida». Noi apprendiamo come Luzhin abbia talmente saturato la sua mente con gli scacchi da vedere tutta la sua vita in termini scacchistici. Ecco la sua goffa corte alla donna che diventerà sua moglie:

Luzhin iniziò con una serie di mosse caute, di cui egli stesso percepiva solo vagamente il significato, la sua singolare dichiarazione d'amore. «Vai avanti, dimmi di più», ella ripeteva, malgrado avesse notato in che silenzio cupo e ottuso egli fosse sprofondato.

Egli si sedette appoggiandosi sul suo bastone da passeggio, pensando che con una mossa di cavallo di quel cedro piantato sul pendio assolato si sarebbe potuto prendere quel palo del telegrafo laggiù, e contemporaneamente cercando di ricordarsi di cosa esattamente aveva appena parlato (p. 97).

Con una spalla appoggiata contro il suo petto, ella cercò con un dito di sollevargli delicatamente le palpebre e la leggera pressione sul suo bulbo oculare provocò un strano lampo di luce nera, un lampo come quello del suo cavallo nero che si sarebbe avventato sul pedone se Turati lo avesse fatto uscire alla settima mossa, come aveva fatto nel loro ultimo incontro. (p. 114)

Ecco uno scorcio dello stato della sua mente dopo il crollo:

Egli si ritrovò in un edificio fumoso dove siedevano rumorosi fantasmi. In ogni angolo si stava svolgendo un attacco – e spingendo via i tavoli,

un secchio da cui sporgeva un pedone di vetro dal collo dorato e un tamburo suonato da un cavaliere su un cavallo dalla folta criniera, egli si aprì la strada verso una lenta porta girevole [...] (p. 139).

Questi temi sono «immagini» sotto molti aspetti, perché gli scacchi sono un gioco spaziale e anche l'identità dei pezzi è usualmente fissata dalle loro forme. Ma il potere che gli scacchi hanno sulla mente di Luzhin non si limita alle proprietà visive o spaziali degli scacchi – qualsiasi cosa possa essere catturata da fotografie o film di una scacchiera, i suoi pezzi in movimento. In realtà, queste proprietà visive contribuiscono solo all'aspetto più superficiale della sua immaginazione. Molto più potente è la *disciplina* offerta dalle regole e la tattica del gioco; è con l'astratta struttura degli scacchi che egli è diventato così ossessivamente familiare e sono le sue abitudini d'indagine su questa struttura che conducono la sua mente di «pensiero» in «pensiero».

[Egli] avrebbe subito notato con disperazione di essere stato di nuovo imprudente e che una mossa delicata nella sua vita era stata appena fatta, continuando impietosamente la combinazione fatale. Allora egli avrebbe deciso di raddoppiare la sua attenzione e di prendere nota di ogni secondo della sua vita, perché le trappole possono essere ovunque. Ed egli era angosciato soprattutto dall'impossibilità di escogitare una difesa razionale, giacché l'intento del suo avversario era ancora oscuro. (p. 227)

Quando impari ad andare in bicicletta o a guidare l'automobile, incontri una nuova struttura di azioni possibili, con nuovi limiti, segnali, abitudini, prospettive, una specie di astratto labirinto comportamentale in cui impari rapidamente a muoverti. Diventa presto una «seconda natura»; incorpori rapidamente la struttura di quel fenomeno esterno nella tua personale struttura di controllo. Nel corso di tale processo puoi cadere in fasi di indagine ossessiva, quando non puoi distogliere la mente dalle nuove mosse. Mi ricordo di un breve periodo di mania per il bridge durante la mia adolescenza in cui avevo sogni ossessivi e insensati sul bridge. Sognavo di fare la stessa astuzia centinaia di volte o di «dichiarare» durante le conversazioni con i miei professori e compagni. Le mie fantasticherie ipnagogiche (quei momenti quasi allucinatori che la gente ha qualche volta quando si sta per addormentare o si sta per svegliare) erano colme di problemi come «qual è la risposta

corretta ad un'apertura preventiva di tre libri – quattro coltelli o quattro forchette?».

Quando nel mondo si incontrano nuove strutture astratte – la notazione musicale, un linguaggio di programmazione per calcolatori, la giurisprudenza, il gioco del baseball – è piuttosto comune ritrovarsi a percorrere avanti e indietro le sue strade specifiche, a costruirsi le proprie strategie – sprofondandovi realmente e cercando di sentirsi a casa. Luzhin è solo un caso estremo; egli ha solo una struttura con cui giocare e la usa per ogni cosa; e va a finire che essa domina tutte le altre sue strutture mentali, convogliando i suoi pensieri in una struttura così rigida da sembrare una sequenza di istruzioni in un programma di una macchina di von Neumann.

Si pensi a tutte le strutture che impariamo, a scuola o altrove: leggere l'ora, l'aritmetica, il denaro, le linee degli autobus, usare il telefono, ecc. Ma tra tutte le strutture con cui diventiamo familiari nel corso della nostra vita, certamente la fonte di disciplina più potente e pervasiva per la nostra mente è la nostra lingua materna. (Spesso si capisce meglio per contrasto; Oliver Sacks, in *Vedere voci*, 1989, attira vividamente l'attenzione sulle ricchezze che il linguaggio apporta alla mente, raccontando la terribile povertà in cui si trova la mente di un bambino sordo se è stato privato di un pronto accesso a un linguaggio *naturale* – segni o linguaggio di segni.) Nel capitolo 8 abbiamo visto come lo stesso vocabolario a nostra disposizione influenza non solo il modo in cui parliamo agli altri, ma anche il modo in cui parliamo a noi stessi. Oltre al contributo *lessicale* c'è quello *grammaticale*. Come Levelt osserva (1989, sez. 3.6), le strutture obbligatorie di frasi nelle nostre lingue sono come tante guide a nostro fianco, che ci ricordano di controllare questo, badare a quello, ci *richiedono* di organizzare fatti in una certa maniera. Alcune di queste strutture potrebbero essere innate, come Chomsky e altri hanno sostenuto, ma non è davvero importante dove tracciare la linea che divide le strutture che sono geneticamente depositate nel cervello e quelle che vi entrano come memi. Queste strutture, reali o virtuali, stendono i binari sui quali i «pensieri» possono poi viaggiare.

Il linguaggio infetta i nostri pensieri e ne altera la flessione a ogni livello. Le parole nel nostro vocabolario sono dei catalizzatori che possono far precipitare contenuti specifici quando una parte del cervello prova a comunicare con un'altra. Le strutture grammaticali impongono una disciplina sulle nostre abitudini di pensiero, forgiando i modi in cui scandagliamo i nostri *data*

base, tentando, come l'avicoltore platonico, di far venire l'uccello giusto quando chiamiamo. Le strutture delle storie che apprendiamo ci guidano ad un livello differente, suggerendoci di porci le domande che sono verosimilmente più importanti nelle nostre attuali circostanze.

Nulla di tutto ciò ha il minimo senso se continuiamo a pensare alla mente come idealmente razionale e perfettamente unificata e trasparente a se stessa. A che serve *parlare a se stessi*, se si conosce già cosa si intende dire? Ma una volta che intravediamo la possibilità di una comprensione parziale, di una razionalità imperfetta, di una problematica comunicazione tra parti, possiamo capire come le potenti forze che un linguaggio libera in un cervello possono essere sfruttate in varie forme di *bootstrapping*, alcune benefiche e altre malefiche.

Ecco un esempio:

Sei magnifico!

Eccone un altro:

Sei patetico!

Tu sai cosa significano queste frasi. Sai anche che le ho introdotte inaspettatamente solo per sorreggere un argomento filosofico e che non costituiscono un atto linguistico rivolto a nessuno. Certamente non ti sto né adulando né insultando, e non c'è nessun altro in giro. Ma non potresti adularti, o insultarti, da solo, servendoti di una o l'altra delle mie frasi, ripetendola a te stesso, più volte, «con enfasi»? Prova, se osi. Qualcosa accade. Non credi a te stesso neanche per un attimo (ti dici), ma scopri che dirti quelle parole provoca in effetti qualche reazione, forse anche un qualche arrossamento alle orecchie, assieme a risposte, repliche, dinieghi, immagini, ricordi, progetti. Queste reazioni possono ovviamente essere nei due sensi. Dale Carnegie aveva ragione sul potere del pensiero positivo, ma, come la maggior parte delle tecnologie, il pensare è più facile da creare che da controllare. Quando parli a te stesso, non devi necessariamente credere a te stesso per far scattare le reazioni. Ci sarà sicuramente qualche reazione, e sarà sicuramente connessa in un modo o nell'altro al significato delle parole con cui ti sei stimolato da solo. Una volta che le reazioni sono avviate, possono condurre la tua mente in luoghi dove ti ritrovi a credere

a te stesso, dopo tutto; quindi fa' attenzione a quello che dici a te stesso.

Il filosofo Justin Leiber così riassume il ruolo del linguaggio nel forgiare la nostra vita mentale:

Se consideriamo noi stessi dal punto di vista informatico, non possiamo evitare di osservare che il linguaggio naturale è il nostro più importante «linguaggio di programmazione». Ciò significa che noi comunichiamo e capiamo gran parte delle nostre conoscenze e attività nei termini del nostro linguaggio naturale [...] Si potrebbe dire che il linguaggio naturale sia stato il nostro primo grande artefatto originale e, poiché, come comprendiamo sempre meglio, i linguaggi sono macchine, così il linguaggio naturale, con il nostro cervello su cui farlo girare, è stata la nostra prima invenzione del calcolatore universale. Si potrebbe dire così se non fosse per il malcelato sospetto che il linguaggio non sia qualcosa che abbiamo inventato ma qualcosa che siamo diventati, non qualcosa che abbiamo costruito ma qualcosa in cui noi creiamo e ricreiamo noi stessi. (Leiber, 1991, p. 8.)

L'ipotesi che il linguaggio giochi un ruolo così importante nel pensare potrebbe sembrare a prima vista una versione dell'ipotesi molto discussa che esista un *linguaggio del pensiero*, un unico mezzo in cui si svolge tutta la cognizione (Fodor, 1975). C'è un'importante differenza, comunque. Molto appropriatamente Leiber chiama il linguaggio naturale un linguaggio di programmazione per il cervello, ma noi possiamo distinguere tra i linguaggi di programmazione di alto livello (come il Lisp, il Prolog e il Pascal) e il più basilare «linguaggio macchina» o l'appena meno basilare «linguaggio assemblatore» tramite i quali vengono composti i linguaggi di alto livello. I linguaggi di alto livello sono macchine virtuali e in un calcolatore creano strutture (temporanee) che lo dotano di una configurazione particolare di punti di forza e debolezza. Il prezzo che si paga per rendere alcune cose «facili da dire» è rendere altre «difficili da dire» o perfino impossibili. Una macchina virtuale del genere può strutturare solo una parte delle competenze del calcolatore, lasciando intatte le altre parti del meccanismo sottostante. Tenendo in mente questa distinzione, è plausibile sostenere che i dettagli di un linguaggio naturale – il vocabolario e la grammatica dell'inglese o del cinese o dell'italiano – vincolino un cervello nella maniera dei linguaggi di programmazione di alto livello. Ma c'è una bella distanza fra questo e l'asserire la dubbia ipotesi che un tale linguaggio naturale fornisca la struttura *a tutti i livelli*. In effetti, Fodor ed altri che

hanno difeso l'idea del linguaggio del pensiero insistono in genere nel dire che *non* stanno parlando del livello al quale i linguaggi *umani* esplicano i loro vincoli. Stanno parlando di un livello di rappresentazione più profondo, meno accessibile. Fodor una volta chiarì questo punto con l'aiuto di una divertente confessione: egli riconobbe che quando si trovava nei momenti più difficili dei propri pensieri, l'unico tipo di espressione linguistica di cui era conscio era qualcosa di simile a «Forza Jerry, puoi farcela!». Quelli potevano essere i suoi «pensieri», e abbiamo già visto come possono giocare effettivamente un ruolo nell'aiutarlo a risolvere i problemi che lo assillavano, ma difficilmente costituiscono il materiale di cui sono fatte le inferenze percettive, le ipotesi da verificare e le altre transazioni postulate al livello basilare del linguaggio del pensiero. Keynes aveva ragione a resistere alla scelta che contrappone le parole alle immagini; i media usati dal cervello sono solo debolmente analoghi ai media rappresentazionali della vita pubblica.

3. *Riferire ed esprimere*

Lentamente, ma con sicurezza, stiamo continuando a scalfire l'idea del Teatro Cartesiano. Nel capitolo 8 abbiamo abbozzato un'alternativa all'Autore Centrale e ora abbiamo appena visto come resistere all'idea di un CRT interno. Temo, però, che siano solo colpi superficiali; il Teatro Cartesiano è ancora in piedi, esercita ancora le sue tenaci attrattive sulla nostra immaginazione. È tempo di cambiare tattica e di attaccarlo dall'interno, facendo esplodere il Teatro Cartesiano mostrando la sua incoerenza nei suoi propri termini. Vediamo cosa succede quando seguiamo la tradizione, quando accettiamo i termini dell'ordinaria «psicologia del senso comune» in modo letterale. Possiamo iniziare riesaminando alcune delle affermazioni plausibili fatte da Otto all'inizio del capitolo 8:

Quando *io* parlo, [diceva Otto] intendo dire ciò che dico. La mia vita cosciente è privata, ma posso decidere di divulgarne alcuni aspetti; posso decidere di dirti varie cose sulla mia esperienza attuale o passata. Quando lo faccio, formulo delle frasi che confeziono attentamente sul materiale che voglio riferire. Posso fare la spola tra l'esperienza e il resoconto che sto per formulare, confrontando le parole con l'esperienza per

essere sicuro di aver trovato *les mots justes*. [...] Mi rivolgo alla mia esperienza cosciente e arrivo ad un giudizio sulle parole che meglio renderebbero giustizia alle sue peculiarità. Quando sono soddisfatto, quando ritengo di aver formulato un resoconto accurato, lo esprimo. Grazie al mio resoconto introspettivo, tu puoi venir a sapere alcune caratteristiche della mia esperienza cosciente.

Alcune di queste affermazioni si adattano perfettamente al modello di produzione del linguaggio che abbiamo proposto nel capitolo 8. Il processo di andirivieni per confezionare le parole adattandole al contenuto dell'esperienza può essere visto come il pandemonio che accoppia i folletti-parola ai folletti-contenuto. Ciò che manca, naturalmente, è l'Io Interno il cui giudizio dirige l'accoppiamento. Ma sebbene Otto seguiti a dire quello che «io scelgo» e quello che «io giudico», l'introspezione non sembra dargli ragione.

Noi abbiamo un accesso limitato ai processi che «ci forniscono» le parole da dire, esse «ci capitano» persino nei casi in cui parliamo deliberatamente, ripetendo silenziosamente i nostri atti linguistici prima di pronunciarli. I candidati per qualcosa da dire spuntano fuori, ma non sappiamo da dove. O ci troviamo già a dirli o ci troviamo a controllarli, talvolta per scartarli, talvolta per modificarli leggermente prima di dirli, ma anche questi occasionali passi intermedi non ci forniscono ulteriori suggerimenti su cosa avviene in noi. Ci troviamo solo nella condizione di accettare o scartare questa o quella parola. Se ci sono ragioni per i nostri giudizi, raramente sono contemplate prima dell'azione, più normalmente sono semplicemente «retrospettivamente ovvie». («Stavo per usare il termine *infecondo* ma mi sono fermato, perché sarebbe apparso pretenzioso».) Quindi non abbiamo davvero nessun accesso privilegiato ai processi che avvengono in noi e che ci portano dal pensiero al discorso. *Potrebbero* essere prodotti da un pandemonio, a quanto ne sappiamo.

Eppure [seguita Otto] il modello Pandemonio lascia fuori del processo un livello o uno stadio. Quello che manca al tuo modello non è una proiezione nello «spazio fenomenologico» di un Teatro Cartesiano – che idea ridicola! – ma un altro livello di articolazione nella psicologia del parlante. Non basta che le parole siano legate insieme da una qualche danza interiore e poi pronunciate; se devono costituire il *resoconto* dello stato mentale

cosciente di qualcuno, devono essere basate in qualche modo su un atto di *apprensione* interna. Quello che il modello Pandemonio lascia fuori è lo stato personale di consapevolezza che guida il discorso.

Indipendentemente dal fatto se abbia o no ragione, Otto esprime certamente l'opinione comune: questo è proprio il modo in cui noi concepiamo la nostra capacità di parlare alla gente dei nostri stati coscienti. In una serie di scritti recenti, il filosofo David Rosenthal (1986, 1989, 1990a, b) ha analizzato questo concetto ordinario di coscienza e la sua relazione con i nostri concetti di *riferire* ed *esprimere*. Egli mette in luce alcune caratteristiche strutturali di cui ci possiamo servire. Primo, possiamo usare la sua analisi per vedere, dall'interno, qual è la prospettiva ordinaria e perché è così stringente. Secondo, possiamo far vedere come discredita l'idea degli zombi – senza

10.7 NORELDO, LA MERAVIGLIA MENTALE, LEGGE LA MENTE DEL SUO GATTO NED.

Drawing by Gahan Wilson © 1990 The New Yorker Magazine Inc. Reprinted with permission

aiuto dall'esterno. Terzo, possiamo rivoltare contro se stessa la prospettiva ordinaria e usare le difficoltà che incontriamo per giustificarne una migliore, una che conservi ciò che c'è di giusto nell'opinione tradizionale ma scarti la sua cornice cartesiana.

Cosa accade quando parliamo? Nel cuore nella nostra concezione ordinaria risiede una banalità: a patto che non stiamo mentendo o siamo insinceri, noi diciamo quello che pensiamo. Per essere un po' più specifici, noi esprimiamo uno dei nostri pensieri o credenze. Supponiamo, per esempio, che io veda il mio gatto aspettare ansiosamente davanti al frigorifero e dica: «Il gatto vuole la sua cena». Questo esprime la mia credenza che il gatto vuole la sua cena. *Esprimendo* la mia credenza, sto *riferendo* quello che ritengo sia un fatto sul gatto. In questo caso sto riferendo il desiderio del gatto di essere nutrito. È importante notare che non sto *riferendo* la mia credenza o *esprimendo* il desiderio del gatto. Il gatto sta esprimendo il suo desiderio sedendo ansiosamente accanto al frigorifero e io, notando questo, lo uso come base – come prova – per il mio resoconto. Ci sono molte altre maniere per esprimere uno stato mentale (come un desiderio), ma un solo modo per riferirne uno, ed è pronunciando un atto linguistico (orale o scritto o altrimenti segnalato).

Uno dei modi più interessanti per *esprimere* uno stato mentale è di *riferire* un altro stato mentale. Nell'esempio, io riferisco il desiderio del gatto, esprimendo così la mia credenza sul desiderio del gatto. Il mio comportamento è una prova non solo del fatto che il gatto ha quel desiderio ma anche che io credo che abbia quel desiderio. Avrei potuto dare, comunque, altre prove differenti della mia credenza – forse alzandomi silenziosamente dalla poltrona e preparando la cena per il gatto. Questo avrebbe *espresso* la stessa credenza senza *riferire* alcunché. O avrei potuto rimanere seduto e sollevare le ciglia, esprimendo *non intenzionalmente* la mia esasperazione per il desiderio del gatto di voler essere nutrito proprio quando mi ero appena comodamente seduto nella mia poltrona. Esprimere uno stato mentale, deliberatamente o no, equivale semplicemente a fare qualcosa che fornisca una prova evidente di quello stato mentale, lo renda manifesto a un altro osservatore – un lettore della mente, se si vuole. Riferire uno stato mentale, invece, è un'attività più sofisticata, sempre intenzionale e che implica il linguaggio.

Ecco, allora, un indizio importante sull'origine del modello del Teatro Cartesiano: la nostra ordinaria psicologia del senso comune tratta il riferire il proprio stato mentale sul modello del

riferire gli eventi del mondo esterno. Il mio resoconto sul gatto che vuole la sua cena è basato sulla mia osservazione del gatto. Il mio resoconto esprime la mia credenza che il gatto voglia la sua cena, una credenza sul desiderio del gatto. Chiamiamo le credenze sulle credenze, i desideri sui desideri, le credenze sui desideri, le speranze sulle paure, e così via stati mentali *del secondo ordine*. E se io (1) *credo* che tu (2) *pensi* che io (3) *voglia* una tazza di caffè, questa mia credenza è una credenza del *terzo* ordine. (Per l'importanza degli stati mentali di ordine superiore nelle teorie della mente, vedi il mio *The Intentional Stance*, 1987a.) Non ci sono dubbi che ci siano differenze importanti e salienti segnate da queste ordinarie distinzioni quando vengono applicate in modo non riflessivo – quando x crede che y sia in qualche stato mentale e $x \neq y$. C'è tutta la differenza che intercorre tra il caso in cui il mio gatto vuole essere nutrito e io lo so e il caso in cui il mio gatto vuole essere nutrito e io non lo so. Ma che dire dei casi riflessivi, quando $x = y$? La psicologia del senso comune li tratta allo stesso modo.

Supponiamo che io riferisca che *io* voglio essere nutrito. Secondo il modello standard dovrei star esprimendo una *credenza del secondo ordine* sul *mio* desiderio. Quando riferisco il mio desiderio, esprimo una credenza del secondo ordine – la mia credenza sul mio desiderio. Che succede se io *riferisco* questa credenza del secondo ordine, dicendo: «Credo che voglio essere nutrito»? Questo resoconto deve esprimere una credenza del *terzo* ordine – la mia credenza che io creda effettivamente nel fatto che voglio essere nutrito. E così via. I nostri concetti ordinari di ciò che è implicito nel parlare sinceramente generano in questo modo una schiera di stati mentali apparentemente separati: il mio desiderio è separato dalla mia credenza di avere il desiderio, che è separato dalla mia credenza di avere la credenza di avere il desiderio, e così via.

La psicologia del senso comune procede anche ad altre distinzioni. Come Rosenthal osserva (assieme a molti altri), essa distingue le *credenze*, che sono stati disposizionali sottostanti, dai *pensieri*, che sono stati occasionali o episodici – eventi passeggeri. La tua *credenza che i cani siano animali* è esistita per anni come uno stato della tua mente, ma il fatto che proprio ora io abbia attirato la tua attenzione su ciò, ha generato un *pensiero* in te – il pensiero che i cani siano animali, un evento che senza dubbio non ti sarebbe capitato ora senza la mia provocazione.

Ne segue, naturalmente, che ci possono essere pensieri del primo ordine e pensieri di ordine superiore – pensieri su pen-

sieri (su pensieri...). Ecco, allora, il passo cruciale: quando io esprimo una credenza – come la mia credenza di voler essere nutrito – non esprimo direttamente la credenza di ordine superiore; ciò che accade è che la mia credenza sottostante provoca un pensiero episodico, il *pensiero* di ordine superiore di voler essere nutrito, e io esprimo *quel pensiero* (se scelgo di farlo). Tutto questo è implicito, sostiene Rosenthal, nel nostro modello di senso comune del *dire quel che si pensa.*

Poiché il tratto distintivo degli stati di coscienza umana è che possono essere riferiti (se escludiamo l'afasia, la paralisi, o l'essere legato e imbavagliato, per esempio), ne segue, secondo l'analisi di Rosenthal, che gli «stati coscienti devono essere accompagnati da appropriati pensieri di ordine superiore e gli stati mentali non coscienti non possono essere così accompagnati» (1990b, p. 16). Il pensiero di ordine superiore in questione deve essere naturalmente sullo stato che accompagna; deve essere il pensiero che si è nello stato di ordine inferiore (o si è appena stati – il tempo passa). Ciò sembra che stia per generare un regresso all'infinito di stati coscienti o pensieri di ordine superiore, ma Rosenthal sostiene che la psicologia del senso comune permette una sorprendente inversione: *il pensiero di secondo ordine non deve esso stesso essere cosciente affinché il suo oggetto del primo ordine sia cosciente.* Puoi esprimere un pensiero senza esserne cosciente, così puoi esprimere un pensiero del *secondo ordine* senza essere cosciente di *esso* – tutto quello di cui devi essere cosciente è il suo oggetto, il pensiero del primo ordine che *riferisci.*

Questo può sembrare sorprendente dapprincipio, ma se ci riflettiamo sopra lo riconosciamo come un fatto familiare in una nuova prospettiva: non ti rivolgi al pensiero che esprimi, ma all'oggetto (o oggetti) *su cui* è il pensiero. Rosenthal prosegue sostenendo che sebbene alcuni pensieri del secondo ordine *siano* coscienti – in virtù di pensieri del terzo ordine su di essi – questi casi sono relativamente rari. Si verificano con i pensieri esplicitamente introspettivi che riferiamo (anche a noi stessi) solo se ci troviamo in uno stato di iper-auto-coscienza. Se ti dico: «Ho un dolore», io riferisco uno stato cosciente, il mio dolore, ed esprimo una credenza del secondo ordine – la mia credenza di avere un dolore. Se, facendomi filosofico, dico: «Penso [o sono sicuro o credo] di avere un dolore», io riferisco così un pensiero del secondo ordine, esprimendo un pensiero del terzo ordine. Normalmente, però, io non avrei un tale pensiero del terzo ordine e quindi non sarei cosciente di tale pensiero del secondo

ordine; lo avrei espresso, dicendo: «Ho un dolore», ma non ne sarei normalmente cosciente.

Quest'idea di pensieri inconsci di ordine superiore potrebbe sembrare dapprincipio eccessiva o paradossale, ma la categoria di episodi in questione non è realmente controversa, anche se il termine «pensiero» non è normalmente usato per riferirsi a loro. Rosenthal usa «pensiero» come un termine tecnico – seguendo approssimativamente la prassi di Cartesio, infatti – per coprire *tutti* gli stati episodici dotati di contenuto e non solo gli episodi che noi chiameremmo normalmente pensieri. Così una fitta di dolore o un fugace turbamento erotico sarebbero considerati pensieri per Cartesio e Rosenthal. A differenza di Cartesio, comunque, Rosenthal insiste sull'esistenza di pensieri inconsci.

I pensieri inconsci sono, per esempio, eventi percettivi inconsci o attivazioni episodiche di credenze, che avvengono naturalmente – che *devono* avvenire – nel corso del normale controllo del comportamento. Supponi di rovesciare una tazza di caffè sul tavolo. In un lampo, salti dalla sedia, evitando per poco il caffè che gocciola lungo il bordo. Non sei stato cosciente di pensare che la superficie del tavolo non avrebbe assorbito il caffè, o che il caffè, un liquido che obbedisce alla legge di gravità, sarebbe scivolato lungo il bordo, ma pensieri inconsci del genere devono essersi verificati – giacché se la tazza avesse contenuto sale o se la superficie del tavolo fosse stata coperta da una tovaglia, non saresti saltato su. Tra tutte le tue credenze – sul caffè, sulla democrazia, sul baseball, sul prezzo del tè in Cina – queste e poche altre sono state immediatamente rilevanti nelle tue circostanze. Se dobbiamo menzionarle nella spiegazione del tuo rapido balzo, devono essere state momentaneamente raggiunte o attivate o in qualche modo intercettate per contribuire al tuo comportamento, ma naturalmente questo è accaduto inconsciamente. Questi episodi inconsci costituirebbero un esempio di ciò che Rosenthal chiama pensieri inconsci. (Abbiamo già incontrato pensieri inconsci in qualche esempio precedente: per esempio, le percezioni inconsce delle vibrazioni nelle dita che permettono di identificare, consciamente, la granulosità superficiale degli oggetti toccati con la bacchetta, il ricordo inconscio della donna con gli occhiali, che condusse all'erronea esperienza della donna che passava velocemente.)

Rosenthal afferma che trovando un modo per definire la coscienza in termini di stati mentali *in*consci (i connessi pensieri di ordine superiore), egli ha scoperto *all'interno della psicologia del senso comune* un modo per porre le fondamenta di una teoria non

misteriosa, non circolare della coscienza (1990b). Ciò che distingue uno stato cosciente da uno stato non cosciente, egli sostiene, non è qualche proprietà intrinseca e inesplicabile, ma la semplice proprietà di avere un pensiero di ordine superiore su di esso che lo accompagna. (Vedi Harnad, 1982, per una strategia simile con qualche variante interessante.) Questa teoria sembra andar bene per la psicologia del senso comune: non si compiace soltanto di rimanere nel mistero; ha le risorse, ben sfruttate da Rosenthal, di articolare una spiegazione del suo concetto primario, la coscienza, in termini dei suoi concetti secondari e meno problematici. Un vantaggio ulteriore, se optiamo per questa analisi, è l'uso che se ne può fare per sgretolare una distinzione ritenuta molto netta tra esseri coscienti e zombi.

4. *Zombi, zimbi e l'illusione dell'utente*

> La mente è una configurazione percepita da una mente. Questo forse è un circolo, ma non è né vizioso né paradossale.
>
> Douglas Hofstadter (1981), p. 197

Gli zombi dei filosofi, come si ricorderà, sembrano capaci di eseguire atti linguistici, sembrano capaci di riferire sui loro stati di coscienza, sembrano capaci di introspezione. Ma in realtà non sono affatto coscienti, malgrado il fatto che siano, nei loro momenti migliori, comportamentalmente indistinguibili da una persona cosciente. Essi possono avere degli stati interni dotati di contenuto funzionale (quel tipo di contenuto che i funzionalisti possono assegnare ai meccanismi interni di un robot), ma si tratta di stati inconsci. Shakey – come lo abbiamo immaginato noi – è un tipico zombi. Quando «riferisce» uno stato interno, quello che viene riferito non è uno stato cosciente, poiché Shakey non ha stati coscienti, ma uno stato inconscio che causa in lui un altro stato inconscio che dirige il processo di generazione ed esecuzione di un cosiddetto atto linguistico composto di formule «preconfezionate» (abbiamo sempre permesso ad Otto di insistere su questo).

Shakey non decide *prima* che cosa *riferire*, dopo aver osservato qualcosa che avviene in lui, e *poi* si preoccupa di come *esprimerlo*; Shakey si ritrova semplicemente con delle cose da dire. Shakey non aveva alcun accesso al perché voleva dire che

tracciava delle linee sui confini di chiaro-scuro delle sue immagini mentali – era semplicemente costruito così. Il punto centrale del capitolo 8, comunque, era che, contrariamente alle apparenze, lo stesso vale per noi. Noi non abbiamo nessun accesso speciale al perché vogliamo dire quello che ci ritroviamo a dire; siamo semplicemente costruiti così. Ma, a differenza di Shakey, noi ricostruiamo noi stessi in continuazione, scopriamo nuove cose che vogliamo dire come risultato della riflessione su quello che abbiamo appena scoperto di aver voluto dire, e così via.

Ma uno Shakey più sofisticato non potrebbe fare altrettanto? Shakey era uno zombi particolarmente rozzo, ma ora possiamo immaginare uno zombi più complesso e più realistico, in grado di osservare le proprie attività, incluse perfino le proprie attività interne, in una spirale indefinitamente crescente di riflessività. Chiamerò una tale entità riflessiva uno *zimbo*. Uno zimbo è uno zombi che, come risultato dell'auto-osservazione, possiede stati informazionali interni (ma inconsci) di ordine superiore su altri stati informazionali di ordine inferiore. (Non fa differenza per questo esperimento ideale se uno zimbo sia considerato un robot o un umano o un marziano.) Quelli che credono che il concetto di zombi sia coerente dovranno sicuramente accettare la possibilità di uno zimbo. Uno zimbo è soltanto uno zombi molto complesso da un punto di vista comportamentale, grazie a un sistema di controllo che permette l'autorappresentazione ricorsiva.

Si consideri ora come potrebbe comportarsi uno zimbo nel test di Turing, la famosa proposta di Alan Turing (1950) per verificare operazionalmente se un calcolatore può pensare. Un calcolatore può pensare, dichiarava Turing, se può battere regolarmente un concorrente umano nel «gioco dell'imitazione»: i due rivali sono tenuti nascosti a un giudice umano ma sono capaci di comunicare con lui inviando messaggi tramite terminali di calcolatori. Il concorrente umano prova semplicemente a convincere il giudice che lui o lei è l'essere umano, mentre il concorrente elettronico tenta di fare lo stesso: prova a convincere il giudice che *esso* è l'essere umano. Se il giudice non può con una certa regolarità individuare il calcolatore, il calcolatore viene ritenuto in grado di pensare. Turing aveva proposto il suo test come qualcosa che mettesse fine alle discussioni; gli sembrava ovvio che esso fosse così mostruosamente difficile da superare che qualunque calcolatore che avesse vinto sarebbe stato ritenuto da tutti un pensatore *incredibilmente* bravo. Egli

pensava di aver messo l'asticella così in alto da soddisfare qualunque scettico. Si sbagliava. Molte persone hanno sostenuto che «superare il test di Turing» non sarebbe una prova sufficiente di intelligenza, e certamente non di coscienza. (Vedi Hofstadter, 1981b; Dennett, 1985a e French, 1991, per analisi dei punti di forza e di debolezza del test di Turing e dei suoi critici.)

Ora, le possibilità di uno zimbo nel test di Turing sarebbero altrettanto buone di quelle di qualunque persona cosciente, poiché i concorrenti non mostrano al giudice null'altro che i loro comportamenti, e in particolare solo il loro comportamento verbale (dattiloscritto). Supponiamo, allora, che tu sia il giudice di un test di Turing e che gli atti linguistici (apparenti) di uno zimbo ti abbiano convinto che è cosciente. Ma questi atti linguistici apparenti non avrebbero dovuto convincerti – *ex hypothesi*, perché hai a che fare solo con uno zimbo e gli zimbi non sono coscienti. Avrebbe potuto convincere se stesso, però? Quando uno zimbo rilascia un resoconto, esprimendo il suo stato inconscio del secondo ordine, non c'è nulla che gli impedisca di riflettere (inconsciamente) su questo stato di cose. In effetti, se vorrà essere convincente, dovrà essere in grado di reagire adeguatamente a (o prendere cognizione di) le proprie «affermazioni» rivolte a te.

Supponiamo, per esempio, che lo zimbo sia uno Shakey più sofisticato e che tu, come giudice, gli abbia chiesto di risolvere un problema con l'occhio della sua mente e di spiegare poi come ha fatto. Esso riflette su quanto ti ha detto, secondo cui ha risolto il problema tracciando delle linee su un'immagine mentale. Esso «saprebbe» che quello era ciò che aveva voluto dire e, se riflettesse ulteriormente, probabilmente «saprebbe» che non sapeva perché quello era ciò che voleva dire. Più gli fai domande su ciò che sapeva e ciò che non sapeva su ciò che stava facendo, più diventa riflessivo. Ora, sembra che quello che siamo riusciti ad immaginare sia un essere inconscio che tuttavia ha la capacità di avere pensieri di ordine superiore. Ma, secondo Rosenthal, quando uno stato mentale è effettivamente accompagnato da un pensiero conscio *o* inconscio di ordine superiore, questo *ipso facto* garantisce che lo stato mentale sia uno stato cosciente! Il nostro esperimento ideale discredita l'analisi di Rosenthal o discredita la definizione di zimbo?

Possiamo vedere facilmente che come minimo lo zimbo crederebbe (inconsciamente) di essere in vari stati mentali – precisamente quegli stati mentali che sarebbe in grado di rife-

rire se glielo chiedessimo. *Esso* crederebbe di essere cosciente, anche se non lo fosse! Ogni entità in grado di superare il test di Turing avrebbe la sensazione (errata?) di essere cosciente. In altre parole, sarebbe vittima di un'illusione (cfr. Harnad, 1982). Che tipo di illusione? Un'illusione dell'Utente, naturalmente. Sarebbe «vittima» della benigna illusione dell'utente della propria macchina virtuale!

Ma questo non è un gioco di specchi, una qualche illegittima prestidigitazione filosofica? Come potrebbe esserci un'illusione dell'Utente senza un Teatro Cartesiano in cui perpetrare l'illusione? Sembra proprio che io corra il rischio di essere sopraffatto dalle mie stesse metafore. Il problema è che un'illusione dell'utente di una macchina virtuale *è* ottenuta tramite una presentazione di materiale in una sorta di teatro alla presenza di un pubblico esterno e indipendente, l'Utente, a beneficio del quale lo spettacolo è messo in scena. In questo preciso istante io sto usando un calcolatore, scrivendo queste parole in un *file*, assistito con discrezione da un programma di videoscrittura. Quando interagisco con il calcolatore, ho un accesso limitato agli eventi che si verificano al suo interno. Grazie agli schemi di presentazione escogitati dai programmatori, mi viene offerta un'elaborata metafora audiovisiva, una rappresentazione teatrale interattiva che si svolge sul palcoscenico della tastiera, del mouse e dello schermo. Io, l'Utente, sono soggetto a una serie di illusioni benigne: sembro in grado di muovere il cursore (un aiutante potente e visibile) proprio sul posto nel calcolatore dove conservo il mio *file*, e una volta visto che il cursore è arrivato «lì», premendo un tasto gli faccio richiamare il *file*, che si espande in un lunghissimo foglio che si srotola di fronte a una finestra (lo schermo), a mio volere. Scrivendo vari comandi, premendo vari tasti, posso far accadere nel mio calcolatore un'infinità di cose, e non sono tenuto a conoscere i dettagli; io mantengo il controllo affidandomi alla mia comprensione della dettagliata metafora audiovisiva offerta dall'illusione dell'Utente.

La maggior parte degli utenti possono apprezzare ciò che si verifica all'interno dei calcolatori solo in termini di queste metafore. Questo è uno degli aspetti che rendono così attraente l'analogia tra una macchina virtuale e la coscienza: il nostro accesso a quello che avviene all'interno del nostro cervello ci è sempre apparso limitato; *noi* non dobbiamo sapere come il meccanismo nelle quinte del nostro cervello realizza la sua magia; *noi* abbiamo familiarità con le sue operazioni solo nei

costumi in cui ci si presentano nella metafora interattiva della fenomenologia. Ma se, quando ci serviamo di questa allettante analogia, manteniamo l'«ovvia» separazione tra la Presentazione da una parte e l'Apprezzamento dello Spettacolo da parte dell'Utente dall'altra, sembra che noi finiamo di nuovo proprio nel Teatro Cartesiano. Come potrebbe esserci un'illusione dell'Utente senza questa separazione?

Non potrebbe esserci; l'utente che fornisce la prospettiva grazie alla quale la macchina virtuale diventa «visibile» deve essere una sorta di osservatore esterno – un *Vorsetzer*. E si potrebbe dapprincipio pensare che il concetto di un tale osservatore debba essere il concetto di un osservatore cosciente, ma abbiamo già visto che non è così. Il *Vorsetzer* che sedeva di fronte al sistema CAD nell'originale sistema CADBLIND Mark I non era cosciente, ma era tuttavia limitato nel suo accesso ai lavorii interni del sistema CAD esattamente come un qualunque utente cosciente. E una volta scartati lo schermo e la telecamera inutili, la Presentazione e l'Apprezzamento dell'Utente svaniscono, rimpiazzati, come tante altre volte nella nostra spiegazione, da una schiera di transazioni più modeste. L'«osservatore esterno» può essere incorporato gradualmente nel sistema, lasciandosi dietro poche tracce fossili: porzioni di «interfacciamento» i cui vari formati continuano a limitare il tipo di domande che possono ricevere una risposta e limitando quindi i contenuti che possono essere espressi.[8] Non deve esserci un *unico* posto dove avviene la Presentazione.[9] E come l'analisi di Rosenthal ci suggerisce, perfino il nostro ordinario concetto di coscienza, ancorato alle intuizioni del senso comune o della psicologia popolare, può tollerare la non coscienza degli stati di ordine superiore la cui presenza nel sistema spiega la coscienza di alcuni dei suoi stati.

Ma allora, il processo di riflessione inconscia è la strada che può portare uno zombi a diventare uno zimbo, e *perciò* a rendersi cosciente? Se è così, allora gli zombi devono essere coscienti dopo tutto. Tutti gli zombi sono capaci di pronunciare «atti linguistici» convincenti (si ricordi che sono indistinguibili dai nostri migliori amici) e questa capacità sarebbe magica se i processi o le strutture di controllo che ne sono causalmente responsabili nel cervello (o calcolatore o qualunque cosa) dello zombi non riflettessero sugli atti e i loro contenuti (apparenti o funzionali). Uno zombi potrebbe iniziare la sua carriera in uno stato non comunicativo e non riflessivo, e quindi essere un vero zombi, un essere inconscio, ma non appena inizia a «comuni-

care» con gli altri e con se stesso, diventerebbe dotato di quegli stessi tipi di stati che secondo l'analisi di Rosenthal sarebbero sufficienti per la coscienza.

Se, d'altra parte, l'analisi di Rosenthal della coscienza in termini di pensieri di ordine superiore viene rifiutata, allora gli zombi possono continuare a fare da cavie per gli esperimenti ideali del futuro. Io presento questa parabola degli zombi un po' ironicamente, poiché penso che *né* il concetto di zombi *né* le categorie della psicologia del senso comune sui pensieri di ordine superiore possano sopravvivere se non come vestigia di un logoro credo. Rosenthal comunque ci ha reso un eccellente servizio esponendo la logica di questi concetti ordinari e grazie alla chiara visione che ora abbiamo di loro, possiamo capire quali sarebbero i rimpiazzi migliori.

5. *Problemi con la psicologia del senso comune*

La psicologia del senso comune, nell'analisi di Rosenthal, postula una gerarchia indefinitamente espandibile di pensieri di ordine superiore, concepiti come episodi dotati di contenuto, che sono indipendenti e salienti, e avvengono in tempo reale nella mente. Ma una visione del genere resiste davvero a una seria verifica? Ci sono nel cervello tali stati ed eventi distinti? Se interpretiamo questa idea in senso lato, la risposta deve essere sì. Ci sono certamente alcune familiari differenze psicologiche che possono essere – e in genere sono – descritte in questi termini.

Improvvisamente Dorothy capì che voleva partire – e che era già molto tempo che voleva partire.

In questo caso sembra che Dorothy abbia acquisito una credenza del secondo ordine – avendo un pensiero del secondo ordine – sul suo desiderio solo un certo tempo dopo che questo desiderio era diventato effettivo. Ci sono molti casi quotidiani di questo tipo: «E poi si accorse che stava guardando proprio le chiavi che aveva perso»; «Egli la ama – solo non se ne rende ancora conto». Difficilmente si può negare che queste frasi ordinarie alludano ad autentiche transizioni tra diversi «stati della mente». Ed intuitivamente, come indica Rosenthal, la transizione implica un *diventare cosciente* di uno stato del primo ordine. Quando Freud, generalizzando questi casi quotidiani,

postula un ampio ed arcano reame di stati mentali inconsci, si riferisce proprio a stati in cui i soggetti non credono di essere. Questa gente si trova in stati mentali in cui non ha ancora capito – tramite pensieri di ordine superiore – di essere.

Tale modo di descrivere queste differenze è familiare, ma se sia anche un modo perspicuo è un'altra faccenda. Queste sono tutte transizioni verso uno stato «meglio informato» (per dirla nel modo più neutrale possibile) ed essere meglio informato in questa maniera è in effetti una condizione necessaria per *riferire* (opposto al mero *esprimere*) il precedente «stato mentale». Ora, il modo incauto di descrivere questa situazione sarebbe: *allo scopo di* riferire uno stato o un evento mentale, devi avere un pensiero di ordine superiore che esprimi. Il che fa emergere questa immagine: prima *osservi* (con qualche organo di senso interiore) lo stato o evento mentale, producendo *a causa di ciò* uno stato di *credenza*, il cui avvio è marcato da un *pensiero*, che poi esprimi. Questa catena causale, come abbiamo visto, imita la catena causale del riferire gli eventi esterni ordinari: prima osservi l'evento con l'aiuto degli organi di senso, provocando in te una credenza, e poi un pensiero, che esprimi in un resoconto.

Questo ipotizzato pensiero di ordine superiore è, io penso, «l'altro livello di articolazione» che Otto pensava di poter discernere nella propria psicologia; è il pensiero che le parole di Otto *esprimono* quando egli *riferisce* la propria esperienza cosciente. Ma rispetto al modello di generazione del discorso che abbiamo abbozzato nel capitolo 8, il modello di Otto ha un rapporto causale affatto invertito. Non è che *prima* si va in uno stato di osservazione di ordine superiore, creando un pensiero di ordine superiore, di modo che poi si possa riferire il pensiero di ordine inferiore esprimendo il pensiero di ordine superiore. È piuttosto lo stato del secondo ordine (quello meglio informato) che viene *creato* dallo stesso processo di elaborazione del resoconto. Noi non apprendiamo *prima* la nostra esperienza nel Teatro Cartesiano e *poi*, sulla base di questa conoscenza acquisita, abbiamo la capacità di elaborare un resoconto da esprimere; la nostra *capacità di dire* che cosa si prova *è la base delle* nostre «credenze di ordine superiore».[10]

A prima vista un processo in stile pandemonio di progettazione di atti linguistici appare errato perché sembra lasciar fuori l'osservatore/decisore centrale il cui pensiero deve alla fine essere espresso. Ma questa non è la debolezza di questo modello, è la sua forza. L'emergere dell'espressione è precisamente ciò che crea o fissa il contenuto del pensiero di ordine

superiore espresso. Non è necessario alcun «pensiero» episodico *addizionale*. Lo stato di ordine superiore dipende letteralmente – dipende causalmente – dall'espressione dell'atto linguistico. Ma non necessariamente dall'espressione pubblica di un atto linguistico manifesto. Nel capitolo 7 abbiamo visto come la necessità propria dell'organismo di migliorare sempre più le comunicazioni interne dell'informazione possa aver condotto alla creazione di abitudini di automanipolazione che possono aver preso il posto di un processo evoluzionisticamente più laborioso di creazione di un occhio interno, un reale organo interno con cui tenere sotto controllo il cervello. L'unico modo che un cervello umano ha di arrivare a qualcosa di simile ad uno stato di credenza di ordine superiore, abbiamo sostenuto, è quello di impegnarsi in un processo simile al riferire a se stesso stati del primo ordine.

Dobbiamo abbandonare l'abitudine di postulare osservatori sempre più centrali. Come punto d'appoggio transitorio, possiamo riconcepire il processo non come conoscenza sulla base dell'osservazione ma sul modello della *diceria*. Io credo *p* perché una fonte affidabile mi ha detto *p*. Chi? Io stesso, o in ogni modo, uno o più di uno dei miei «agenti». Questo non è un modo completamente estraneo di pensare; noi parliamo, dopo tutto, della *testimonianza* dei nostri sensi, una metafora che suggerisce che i nostri sensi non vengono «in aula» a *mostrarci* gli indizi, ma piuttosto *ci dicono* varie cose. Affidandoci a questa metafora (finché non ci abituiamo alle complessità di un'alternativa migliore), potremmo basarci su questo motto:

Se non potessi parlare a me stesso, non avrei modo di conoscere quello che stavo pensando.

Questo, per molti aspetti, non è ancora il modo giusto di pensare a ciò. Prima di tutto, c'è una differenza – che stavo oscurando – tra un'entità «che parla a se stessa» e una varietà di sottosistemi «che parlano tra loro». Un'adeguata transizione tra queste due idee verrà negoziata nel capitolo 13, sul sé. Poi, come abbiamo visto, l'enfasi sull'espressione linguistica è esagerata; ci sono altre strategie di automanipolazione e autoespressione che non sono verbali.

Potrà benissimo sembrare che io stia proponendo un misero affare: abbandonare la relativa nitidezza e chiarezza del modello standard della psicologia del senso comune, con la sua gerarchia di osservazioni interne, per un'alternativa approssi-

mativa che non possiamo ancora concepire chiaramente. Ma la chiarezza del modello tradizionale è illusoria, per le ragioni viste al capitolo 5, quando abbiamo esaminato lo strano argomento del *reale sembrare*. Ora possiamo diagnosticare i problemi più precisamente. Otto è il portavoce della psicologia del senso comune e se lo lasciamo continuare si troverà presto impigliato in mille nodi. La posizione di Otto, che si accanisce a estendere «fino in fondo» le categorie della psicologia del senso comune, genera un'esplosione di «stati rappresentazionali» distinti, le cui relazioni reciproche generano enigmi artificiali. Otto prosegue:

Il mio resoconto pubblico di un stato cosciente, se dovessi scegliere di farlo, potrebbe contenere un errore. Potrei fare un lapsus o sbagliarmi sul significato di una parola e quindi informarti male inavvertitamente. Qualunque errore espressivo del genere di cui io non mi accorga potrebbe indurti a credere erroneamente qualcosa su di me – su come stanno *realmente* le cose in me. E il semplice fatto che non mi sia accorto dell'errore non significa che non ci sia un errore. *Da una parte*, c'è la verità su come stanno le cose in me, e poi, *dall'altra*, c'è quello che alla fine dico su come stanno le cose in me (se scelgo di farlo). Sebbene io tenda ad essere una fonte altamente attendibile su ciò che riferisco, c'è sempre spazio per degli errori.

Questa è una di quelle situazioni in cui due mani non bastano. Perché, come Rosenthal ci ha mostrato, oltre a «come stanno le cose in me» e «quello che alla fine dico» sembra che debba esserci un terzo fatto intermedio: la mia credenza su come stanno le cose in me.[11] Perché, quando dico sinceramente quello che dico, intendendo quello che intendo, io esprimo una delle mie credenze – la mia credenza su come stanno le cose in me. In effetti, c'è un quarto fatto intermedio: il mio pensiero episodico che le cose stanno così in me.

La mia credenza su come stanno le cose in me potrebbe essere erronea? O potrebbe essere che io *penso soltanto* che le cose stanno così in me? O in altre parole, potrebbe soltanto *sembrarmi* che questa sia la mia attuale esperienza? Otto voleva una separazione, ma ora siamo minacciati da ben altro: abbiamo una divisione tra l'esperienza soggettiva e la credenza su di essa, tra la credenza e il pensiero episodico che sorge sulla strada dell'espressione verbale, e tra il pensiero e la definitiva espressione. E, come la moltiplicazione delle scope nell'appren-

dista stregone, ci sono altre divisioni in vista una volta accettate queste. Si supponga che io abbia la mia esperienza soggettiva (che è una cosa) e che essa mi fornisca le basi per credere di averla (questa è la seconda cosa) che a sua volta genera il pensiero associato (una terza cosa) che poi mi incita ad un'intenzione comunicativa da esprimere (una quarta cosa), che conduce, finalmente, all'effettiva espressione (una quinta cosa). Ma non c'è spazio per errori che si possano insinuare nella transizione tra le varie cose? Non potrebbe essere che io creda una proposizione, ma, a causa di una erronea transizione tra stati, finisca per *pensare* a una proposizione differente? (Se si può «parlare erroneamente», si può anche «pensare erroneamente»?) Sarebbe possibile elaborare l'intenzione di esprimere una proposizione piuttosto differente da quella a cui si pensa? E non potrebbe una memoria difettosa nel sottosistema delle intenzioni comunicative condurci a iniziare con un messaggio preverbale da esprimere e finire con un differente messaggio preverbale che serve come standard su cui correggere gli errori? Tra ogni paio di cose distinte c'è uno spazio logico per l'errore, e, man mano che moltiplichiamo stati individuali con contenuti definiti, noi scopriamo – o creiamo – molteplici fonti di errore.

C'è una forte tentazione di semplificare questo pasticcio dichiarando che *il mio pensiero (o credenza) su come mi sembra* è equivalente a *come la mia esperienza è effettivamente*. In altre parole, c'è una tentazione a insistere che logicamente non c'è spazio per errori tra loro, poiché sono la stessa cosa. Una tale affermazione ha alcune proprietà positive. Arresta l'esplosione minacciata al primo passo – di solito il punto giusto dove fermare un'esplosione o un regresso – ed esercita qualche autentica attrazione sulla nostra intuizione, il che viene evidenziato adeguatamente dalla domanda retorica: «Che senso plausibile potrebbe essere attribuito all'affermazione che qualcosa solamente mi sembrava di sembrarmi (di sembrarmi...) essere un cavallo?».

Ma qui dobbiamo procedere cautamente, saltellando attorno alle ossa di teorie filosofiche defunte (inclusa qualcuna delle mie – cfr. Dennett, 1969, 1978c, 1979a). Sembrerebbe che potremmo rimanere attaccati alle vecchie e buone categorie della psicologia del senso comune sulle credenze, i pensieri, le credenze sulle credenze, i pensieri sulle esperienze, e simili, ed evitare le perplessità dell'autoconoscenza semplicemente unificando i casi riflessivi di ordine superiore e inferiore: dichiarando che quando io credo di credere *p*, per esempio, ne consegue logicamente che io credo *p*, e nello stesso spirito, quando io

penso di avere un dolore, ne consegue logicamente che io ho un dolore, e così via.

Se le cose stessero così, quando io esprimo una credenza del secondo ordine riferendo una credenza del primo ordine, per esempio, starei in realtà trattando solo con un unico stato, un'unica cosa, e il fatto che nel riferire una cosa io ne stavo esprimendo «un'altra» sarebbe dovuto solo a una mera distinzione verbale, come il fatto che Rossi si era preparato a sposare la sua fidanzata e ha finito con lo sposare sua moglie.

Ma questa unificazione non svolge tutto il lavoro che dovrebbe essere fatto. Per rendersene conto, consideriamo di nuovo il ruolo della memoria, così come viene concepito nella psicologia del senso comune. Anche se è intuitivamente plausibile che non ti puoi sbagliare su come *stiano* le cose in te proprio *adesso*, non è affatto plausibile intuitivamente che tu non ti possa sbagliare su come *stavano* le cose in te *allora*. Se l'esperienza che stai riferendo è un'esperienza passata, nulla esclude che la tua memoria – su cui ti stai basando per il tuo resoconto – possa essere contaminata da un errore. Forse la tua esperienza era davvero di un certo tipo, ma ora la ricordi male e la descrivi in un altro modo. Certamente potrebbe sembrarti *ora* di esserti sembrato *allora* che fosse stato un cavallo – anche se *di fatto* ti sembrò *allora* che fosse una mucca. La possibilità logica di ricordare erroneamente sussiste sempre indipendentemente dalla brevità dell'intervallo temporale tra l'esperienza effettiva e il ricordo successivo – questo è ciò che autorizzava le teorie orwelliane. Ma, come vedemmo nel capitolo 5, l'errore che si insinua nella credenza successiva grazie a una manomissione orwelliana della memoria è indistinguibile – sia dall'esterno che dall'interno – dall'errore che si insinua nell'esperienza originale grazie ad una costruzione staliniana dell'illusione. Così anche se potessimo continuare a dire che hai un accesso «diretto» e «immediato» al tuo attuale *giudizio* (il tuo pensiero del secondo ordine su come le cose ti sembrano ora), tu non sarai per questo capace di escludere la possibilità che esso sia un giudizio erroneo su come ti sembrava un momento fa.

Se individuiamo stati (credenze, stati di coscienza, stati di intenzioni comunicative, ecc.) tramite il loro contenuto – che è il mezzo standard di individuazione nella psicologia del senso comune – finiamo per dover postulare differenze che sono sistematicamente non distinguibili in alcun modo, interno o esterno, e nello stesso tempo perdiamo l'intimità o incorreggibilità soggettiva che si ritiene sia il tratto distintivo della

coscienza. Abbiamo già visto esempi di questo nella discussione al capitolo 5 che contrapponeva i modelli di fenomeni temporali orwelliani a quelli staliniani. E la soluzione non è di favorire una o l'altra dottrina rese possibili dalla psicologia del senso comune, ma di abbandonare questa caratteristica della psicologia del senso comune.

Noi sostituiamo la divisione in stati discreti dotati di contenuto – credenze, meta-credenze, e così via – con un *processo* che serve, nel corso del tempo, ad assicurare un buon accoppiamento tra eventi interni di un'entità che sono portatori di informazione e la capacità dell'entità stessa di *esprimere* nel discorso (alcune delle) informazioni in quegli eventi. Questo è ciò che gli stati di ordine superiore dovevano assicurare (secondo Otto), ma non sono riusciti a penetrare nel loro oggetto rispettandone le articolazioni. Al contrario, hanno postulato articolazioni che erano sistematicamente indiscernibili in natura.

Questi prodotti della psicologia del senso comune continuano a vivere, comunque, come cittadini dei mondi eterofenomenologici dei soggetti le cui visioni del mondo sono effettivamente plasmate da quello schema concettuale. Per dirla tautologicamente, poiché alla gente sembra davvero di avere sia queste credenze sulle esperienze sia le esperienze stesse, queste esperienze e queste credenze sulle esperienze sono entrambe parte di ciò che appare loro. E così dobbiamo spiegare *questo* fatto – non il fatto che le nostre menti *sono* organizzate in gerarchie di «stati» rappresentazionali di ordine superiore di credenze, meta-credenze, e così via, ma che le nostre menti *tendono a sembrarci* così ordinate.

Ho avanzato due ragioni per cercare di spiegare perché tendiamo a trovare questa un'idea così attraente. Primo, perché persistiamo nell'abitudine di postulare un processo di osservazione separato (in questo caso di osservazione *interiore*) che si frappone tra le circostanze che possiamo riferire e il resoconto che rilasciamo – ignorando il fatto che questo regresso di osservatori interni in qualche punto deve essere arrestato da un processo che unifica i contenuti con le loro espressioni verbali senza nessun apprezzatore intermedio del contenuto. E, secondo, perché le comunicazioni interne create in questo modo hanno un notevole effetto nell'organizzare le nostre menti in sistemi riflessivi o autoscrutinanti indefinitamente potenti. Si è affermato spesso, a buona ragione, che tali poteri di riflessione stanno al cuore della coscienza. Noi possiamo usare il modello

estremamente semplificato della psicologia del senso comune come una specie di stampella per l'immaginazione quando proviamo a capire i sistemi che si autoesaminano, ma, quando la usiamo, rischiamo di scivolare nel materialismo cartesiano. Dovremmo iniziare ad imparare a farne a meno, e nel prossimo capitolo faremo qualche altro cauto passo.

/

11

SMANTELLARE IL PROGRAMMA DI PROTEZIONE DEL TESTIMONE

1. *Ricapitolazione*

Nella Parte prima abbiamo passato in rassegna i problemi e formulato alcuni presupposti e alcuni princìpi metodologici. Nella Parte seconda abbiamo abbozzato un nuovo modello della coscienza, il modello delle Molteplici Versioni, e abbiamo cominciato a vedere perché dovrebbe essere preferito al modello tradizionale, il Teatro Cartesiano. Sebbene l'idea del Teatro Cartesiano, una volta resa esplicita, mostri le sue pecche in modo piuttosto drammatico – nessuno si dichiara apertamente materialista cartesiano – le abitudini di pensiero e le presupposizioni implicite che essa ha alimentato continuano a motivare le obiezioni e a distorcere le «intuizioni». Ora nella Parte terza, stiamo indagando le implicazioni del nostro modello alternativo, rispondendo ad una serie di interessanti obiezioni. Alcune di esse tradiscono una fedeltà persistente – sebbene negata – al vecchio caro Teatro Cartesiano.

«Ma dove avviene la comprensione?» Questa domanda si annida al centro della controversia sin dal diciassettesimo secolo. Cartesio incontrò un'ondata di scetticismo quando insistette (correttamente) nell'affermare che i meccanismi cerebrali erano in grado di spiegare almeno gran parte della comprensione. Per esempio Antoine Arnauld, nelle sue obiezioni alle *Meditazioni*, notò che «a prima giunta sembra incredibile che possa accadere, senza il ministerio di nessuna anima, che la luce che si riflette dal corpo del lupo negli occhi della pecora muova i piccoli filamenti dei nervi ottici, e che in virtù di questo movimento, che va sino al cervello, gli spiriti animali siano diffusi nei nervi nella maniera ch'è necessaria perché la pecora prenda la fuga» (1641, p. 197). Cartesio replicò che ciò non era più incredibile della capacità umana di gettare avanti le braccia per proteggersi quando si cade, anche questa una reazione

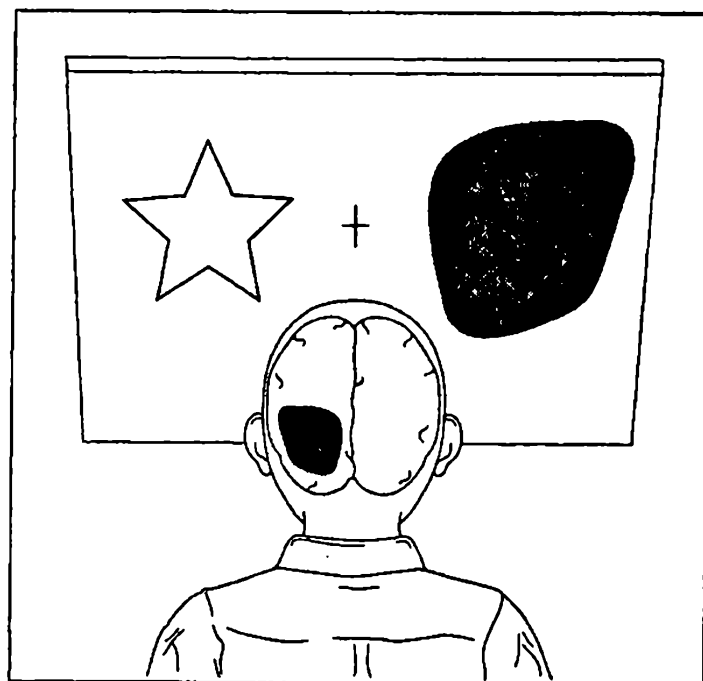
meccanica, eseguita senza l'intervento dell'«anima». Questa idea dell'interpretazione «meccanica» nel cervello è l'intuizione centrale di ogni teoria materialistica della mente, ma cozza contro molte altre concezioni intuitive: la nostra sensazione che, affinché ci sia una *vera* comprensione, è necessaria la presenza di *qualcuno lì dentro* che convalidi i procedimenti, che *testimoni* gli eventi il cui verificarsi costituisce la comprensione. (Il filosofo John Searle sfrutta questa concezione intuitiva nel suo famoso esperimento ideale della Stanza Cinese, che sarà esaminato nel capitolo 14.)

Cartesio era un meccanicista *par excellence* quando si trattava di ogni altro fenomeno della natura, ma quando si trattava della mente umana indietreggiava. Oltre all'interpretazione meccanica, egli sosteneva, il cervello fornisce anche materiale ad un'arena centrale – quello che io ho chiamato il Teatro Cartesiano – dove, negli esseri umani, l'anima può essere un Testimone e pervenire ai propri giudizi. I testimoni hanno bisogno di materiali grezzi su cui basare i propri giudizi. Questi materiali grezzi, siano essi chiamati «dati sensoriali» o «sensazioni» o «sentimenti grezzi» o «proprietà fenomeniche dell'esperienza», sono oggetti scenici senza i quali un Testimone non ha senso. Questi oggetti scenici, tenuti in piedi da varie illusioni, circondano l'idea di un Testimone centrale con una barriera quasi impenetrabile di intuizioni. Il compito di questo capitolo è di provocare una falla in questa barriera.

2. *La visione cieca: una zombità parziale?*

Di tutti i terribili incidenti che si abbattono sulla gente, una minima frazione viene parzialmente riscattata dal fatto di rivelare qualcosa sui misteri della natura agli scienziati che la studiano. Questo è particolarmente vero dei danni cerebrali provocati da traumi (colpi d'arma da fuoco, incidenti stradali e simili), tumori o apoplessia.[1] L'intreccio risultante di disabilità e abilità sopravvissute fornisce talvolta degli indizi importanti – perfino sorprendenti – su come la mente sia realizzata dal cervello. Una delle patologie più sorprendenti, come il suo nome paradossale suggerisce, è la visione cieca; sembra fatta apposta per gli esperimenti ideali dei filosofi: è una disgrazia che trasforma una persona normale e cosciente in uno zombi parziale, un automa inconscio rispetto a certi stimoli, ma una persona normalmente cosciente per il resto. Così non ci sor-

prende che i filosofi abbiano innalzato la visione cieca a una sorta di status mitico come esempio sul quale costruire i propri argomenti. Come vedremo, comunque, la visione cieca, invece di rafforzare il concetto di zombi, lo mina.

Nella visione umana normale, i segnali che giungono agli occhi viaggiano nei nervi ottici attraverso varie stazioni intermedie fino alla *corteccia occipitale* o corteccia visiva, la parte del cervello nel retro del cranio sopra il cervelletto. Le informazioni sul campo visivo sinistro (le due metà sinistre dei campi da ogni occhio) si diffondono sulla corteccia visiva destra, e il campo visivo destro sulla corteccia sinistra. Talvolta un incidente vascolare (per es., la rottura di un vaso sanguigno) può distruggere una porzione di corteccia occipitale, creando un punto cieco o *scotoma*, un buco relativamente largo nell'esperienza visiva del mondo, sul lato opposto al danno.

11.1

Nel caso estremo in cui sia la corteccia visiva destra che quella sinistra sono state distrutte, la persona è resa completamente cieca. Più spesso, l'intera corteccia visiva su un lato del cervello è distrutta da un incidente vascolare, conducendo alla perdita della metà opposta del campo visivo; la perdita della corteccia visiva sinistra produrrà *emianopia destra*, cecità completa nell'emicampo destro.

Che cosa si prova ad avere uno scotoma? Potrebbe sembrare che sia un'esperienza familiare a tutti noi, giacché tutti noi abbiamo un punto cieco nel nostro campo visivo in corrispondenza del luogo nella nostra retina dove non ci sono né baston-

celli né coni perché il nervo ottico esce dal globo oculare in quel punto. Un punto cieco normale, o disco ottico, non è piccolo: oscura un cerchio con un diametro di circa 6 gradi di angolo visivo. Chiudi un occhio e fissa la croce, tenendo la pagina a circa 25 centimetri dagli occhi. Uno dei dischi che cade nel punto cieco scomparirà. Chiudi l'altro occhio, il disco dalla parte opposta scomparirà. (Forse sarà necessario un certo aggiustamento nella distanza della pagina per far avvenire l'effetto. Perché non noti normalmente questa lacuna nel tuo campo visivo? In parte perché hai due occhi e uno supplisce l'altro; i due punti ciechi non si sovrappongono. Ma anche con un occhio chiuso non noterai il tuo punto cieco in moltissime circostanze. Perché? Poiché il tuo cervello non ha mai dovuto trattare segnali in ingresso da questa area della retina, non ha destinato alcuna risorsa per il loro trattamento. Non ci sono homunculi responsabili di ricevere resoconti da quest'area, così quando non arriva nessun resoconto, nessuno si lamenta. Un'assenza di informazione non è la stessa cosa di un'informazione su un'assenza. Affinché tu possa vedere un buco, qualcosa nel tuo cervello deve rispondere ad un contrasto: *o* tra il bordo interno ed esterno – e il tuo cervello non ha alcun meccanismo per far questo in quel punto – *o* tra il prima e il dopo: ora vedi il disco, ora non lo vedi. (In questo modo il disco nero che scompare della figura 11.2 ti avverte del tuo punto cieco.)

11.2

Come i nostri punti ciechi normali, gli scotomi hanno delle posizioni definite e alcuni hanno dei confini netti che possono essere facilmente rilevati da uno sperimentatore, usando uno stimolo come un punto luminoso che può essere mosso a piacere nel campo visivo del soggetto; al soggetto viene chiesto di riferire quando il punto luminoso esce dalla sua esperienza – una variazione dell'esperimento che hai appena fatto per scoprire il tuo punto cieco. I resoconti dei soggetti possono poi essere correlati con mappe del danno corticale, prodotte da analisi del cervello come TAC (tomografia assiale computerizzata) e IRM (immagini di risonanza magnetica). Uno scotoma è differente per un aspetto importante dal punto cieco normale: viene di solito avvertito dal soggetto. Questo non solo perché è più grande di un normale punto cieco; poiché è causato dalla

perdita di cellule nella corteccia visiva che precedentemente «riferivano a» altre cellule nella corteccia che «si occupavano» anche loro dell'informazione da certe regioni della retina, la loro assenza *viene* notata. Le aspettative del cervello sono disattese; qualcosa che dovrebbe esserci manca, una specifica fame epistemica non è soddisfatta. Così il soggetto è normalmente cosciente dello scotoma, ma come una mancanza, non come un'effettiva area nera, come quella che vedresti se qualcuno incollasse un cerchio di carta nera sul tuo parabrezza.

Poiché i normali tragitti visivi nel cervello sono stati interrotti o tagliati, ci si aspetterebbe che i soggetti affetti da scotoma fossero completamente incapaci di raccogliere informazioni dalla vista di ciò che avviene nei loro campi ciechi. Sono ciechi, dopo tutto. E questo è proprio quello che dicono: la loro esperienza visiva all'interno dei confini dello scotoma è nulla – nessun lampo, bordo, colore, scintillio o sfavillio. È la cecità. Ma alcune persone affette da scotoma mostrano un sorprendente talento: malgrado la loro completa mancanza di esperienza visiva cosciente nell'area cieca, possono talvolta «indovinare» con notevole accuratezza se una luce è stata o no appena proiettata nel campo, e anche se si è mostrato un cerchio o un quadrato. Questo fenomeno è chiamato «visione cieca» (Weiskrantz, 1986, 1988, 1990). Come la visione cieca debba essere esattamente spiegata è ancora controverso, ma nessun ricercatore pensa di trovarsi di fronte a un fenomeno «paranormale». Ci sono almeno dieci differenti percorsi tra la retina e il resto del cervello, così, se anche la corteccia occipitale è distrutta, ci sono ancora numerosi canali di comunicazione tramite i quali le informazioni possono andare dalle retine perfettamente normali ad altre aree cerebrali. Recentemente sono stati realizzati molti test su pazienti affetti da visione cieca e non ci sono dubbi che essi riescano a indovinare molto meglio che a caso varie forme semplici, la direzione del movimento, la presenza o l'assenza della luce (in certe condizioni i successi raggiungono addirittura una percentuale del 100 per cento). Nessuna persona con visione cieca ha dimostrato finora la capacità di discriminare i colori nel campo cieco, ma una ricerca recente di Stoering e Cowey (1990) fornisce prove che ciò sia possibile.

Cosa avviene nella visione cieca? Si tratta, come alcuni filosofi e psicologi hanno caldeggiato, di percezione visiva senza coscienza – del tipo che potrebbe mostrare un mero automa? Fornisce una refutazione delle teorie funzionalistiche della

mente (o perlomeno le mette in seria difficoltà) mostrando un caso in cui tutte le *funzioni* della visione sono ancora presenti, ma tutta la parte succosa della *coscienza* è stata prosciugata? Non fornisce niente del genere. Nella loro fretta d'imbrigliare la visione cieca per portare acqua ai loro mulini ideologici, i filosofi hanno talvolta trascurato alcuni fatti piuttosto elementari sul fenomeno della visione cieca e sulla situazione sperimentale in cui può emergere.

Come le «anomalie temporali» analizzate nei capitoli 5 e 6, il fenomeno della visione cieca appare solo quando trattiamo i soggetti dal punto di vista dell'eterofenomenologia. Gli esperimenti non potrebbero essere condotti se gli sperimentatori non potessero dare istruzioni verbali ai soggetti (e avere fiducia di essere stati capiti) e le risposte dei soggetti danno prova della presenza di un sorprendente fenomeno *solo* se sono interpretate come atti linguistici. Ciò è fin troppo ovvio da notare e perciò debbo fermarmi per farlo capire bene.

L'interpretazione da dare alla visione cieca è controversa sotto molti aspetti, ma su un punto il consenso è generale: il soggetto con visione cieca viene in qualche modo informato su un evento nel mondo attraverso gli occhi (questa è la parte «visione»), malgrado non abbia un'esperienza visiva cosciente dell'evento in questione (questa è la parte «cieca»). Più succintamente, la visione cieca implica (1) la ricezione di informazioni visive che (2) sono tuttavia inconsce. La dimostrazione di (1) è diretta: il soggetto riesce molto meglio che a caso in test che scandagliano quell'informazione. La dimostrazione di (2) è più indiretta e indiziaria: i soggetti negano di essere coscienti di tali eventi, e le loro negazioni verbali sono sorrette dalle prove neurologiche di danni cerebrali, da una parte, e dalla coerenza delle loro negazioni, dall'altra; quindi gli crediamo![2]

Questo non è un punto banale. Si noti che ciò che ci sorprende sulla visione cieca svanirebbe immediatamente se potessimo concludere che i soggetti stanno simulando – solamente fingendo di non essere coscienti. O, per esaminare un caso più vicino a quello in questione, si confronti come accettiamo le affermazioni di cecità dei soggetti con visione cieca allo scetticismo con cui accogliamo le stesse affermazioni quando provengono da persone a cui sia stata diagnosticata una «cecità isterica». Talvolta alcune persone che sembrano avere occhi e cervello in perfetto buono stato, per quanto possano determinare i fisiologi, si lamentano di essere ciechi; e sostengono quest'affermazione comportandosi «proprio come un cieco».

Spesso si può trovare una ragione abbastanza plausibile per giustificare perché una persona del genere voglia «diventare» cieca – o per punire se stessa o qualcun altro che deve ora accudirla e addolorarsi, o per negare qualche terribile memoria visiva, o come una specie di risposta di panico a qualche altra malattia o disabilità – così questa è cecità «psicosomatica», se di cecità si tratta. Questa gente è davvero cieca? Potrebbe esserlo. Dopo tutto, si potrebbe sostenere, se il dolore psicosomatico è dolore reale, e la nausea psicosomatica è abbastanza reale da far vomitare, perché la cecità psicosomatica non dovrebbe essere cecità reale?

Le persone affette da cecità isterica *sostengono* di essere cieche, ma, come i soggetti con visione cieca, danno tuttavia delle prove inconfutabili di *recepire informazioni visive*. Per esempio è tipico delle persone affette da cecità isterica dare dei risultati peggiori del caso se si chiede loro di indovinare le caratteristiche visive di un oggetto! Questo è un segno sicuro che stanno in qualche modo utilizzando le informazioni visive per guidare il loro comportamento verso una preponderanza di «errori». Queste persone hanno una misteriosa abilità a trovare sedie in cui inciampare. Eppure, a differenza dei simulatori completi, quando dicono di non aver esperienze visive, sono sincere – credono realmente a quello che dicono. Noi non dovremmo crederci? Come dovremmo trattare i testi di questi due gruppi differenti di persone quando ci accingiamo ad estrapolare i loro mondi eterofenomenologici?

Ecco una situazione in cui le linee di condotta ultra-caute dell'eterofenomenologia pagano i dividendi. Sia i soggetti con visione cieca sia le persone affette da cecità isterica sono apparentemente sincere nelle loro dichiarazioni di essere inconsapevoli di qualunque cosa avviene nei loro campi ciechi. Così i loro mondi eterofenomenologici sono simili – almeno rispetto ai presunti campi ciechi. Eppure c'è una differenza. Noi abbiamo meno conoscenze sui fondamenti neuroanatomici della cecità isterica di quelle che abbiamo sulla visione cieca, eppure ci sentiamo, intuitivamente, molto più scettici sulle affermazioni di chi è affetto dalla prima.[3] Che cosa ci fa sospettare che queste persone non siano affatto cieche *in realtà*, che in qualche modo e in qualche grado siano anche coscienti dei loro mondi visivi? Le circostanze sospettosamente propizie della loro cecità ci fanno meravigliare, ma oltre a queste prove indiziarie indirette c'è una ragione più semplice: noi dubitiamo della loro pretesa cecità perché, *senza sollecitazione*, le persone affette da cecità

isterica *usano* talvolta le informazioni che i loro occhi forniscono loro in modi sconosciuti ai soggetti con visione cieca.

Nelle situazioni sperimentali allestite per studiare la visione cieca è presente un fattore che si adatta così bene alle nostre presupposizioni ordinarie che quasi nessuno si preoccupa di discuterlo (ma vedi Marcel, 1988; van Gulick, 1989; Carruthers, 1989): i soggetti con visione cieca devono essere sollecitati o invitati a comunicare le loro «congetture». Così lo sperimentatore può aver detto, nelle istruzioni iniziali: «Ogni volta che senti un suono, prova ad indovinare», oppure: «Ogni volta che ti do un colpetto sulla mano, dai la risposta». Senza queste imbeccate, i soggetti semplicemente non rispondono.[4]

Immaginiamo una variazione sul tema per saggiare il valore della differenza che abbiamo diagnosticato tra le due malattie. Supponiamo di incontrare una paziente con pretesa visione cieca che non deve essere imbeccata: «spontaneamente» comunica «congetture» (molto migliori del caso, ma non perfette) ogni volta che qualcosa viene presentato nel suo preteso campo cieco. La facciamo accomodare nel nostro laboratorio ed effettuiamo i normali esami per rilevare il presunto scotoma; ella ci dice ogniqualvolta la luce in movimento scompare dal suo campo visivo, come ogni altro paziente con visione cieca. Ma nello stesso tempo fa dichiarazioni spontanee, non sollecitate, come questa: «Sia ben inteso è solo una congettura, ma non avete proiettato una luce nel mio scotoma proprio adesso?». Ma questo solo dopo che lo abbiamo effettivamente fatto. Ciò sarebbe, come minimo, sospetto, e possiamo dire perché.

In generale, il fatto che i soggetti di un esperimento osservino effettivamente le istruzioni date viene ritenuto un segno evidente che essi sono stati in grado di seguire le istruzioni perché hanno avuto un'*esperienza cosciente* dello stimolo pertinente. Ecco perché la seguente istruzione preparatoria sarebbe giudicata insensata:

Ogni volta che sei *cosciente* dell'accensione della luce, premi il pulsante di sinistra; ogni volta che la luce si accende ma *non* ne sei cosciente, premi quello di destra.

Come diavolo farebbe un soggetto a seguire questa istruzione? Gli si chiederebbe di fare l'impossibile: di regolare il suo comportamento su situazioni che gli sono inaccessibili. Sarebbe come dire: «Alza la mano ogniqualvolta qualcuno ti fa l'occhietto senza che te ne accorgi». Uno sperimentatore non

sentirebbe il bisogno di inserire l'avverbio «consciamente», come in

Ogni volta che senti consciamente un suono, prova ad indovinare.

poiché si presuppone normalmente che non si possono regolare le proprie azioni su esperienze *inconsce*, anche se cose del genere avvengono. Per seguire

Ogni volta che avviene *x*, fai *y*

devi essere cosciente del verificarsi di *x*.

Questa è la nostra presupposizione ordinaria, ma questo edificio di ovvietà ha una crepa. Non abbiamo imparato che molti dei nostri comportamenti sono governati da condizioni che noi rileviamo solo inconsciamente? Si considerino le attività che regolano la nostra temperatura corporea, adattano il nostro metabolismo, immagazzinano e recuperano energia, attivano il nostro sistema immunitario; si considerino attività come il battere le ciglia quando una cosa si avvicina o entra nell'occhio, e perfino macroscopici comportamenti pubblici come camminare (senza cadere) e chinare la testa quando una cosa si avvicina improvvisamente. Tutto questo «comportamento» viene controllato senza alcun aiuto dalla coscienza – come Cartesio ha osservato.

Sembra, allora, che ci siano due tipi di condotte comportamentali: quelle controllate dal pensiero cosciente e quelle controllate da processi «ciechi, meccanici» – proprio come i processi che controllano un ascensore automatico. Se un ascensore automatico deve aderire alla linea di condotta di non portare più di mille chilogrammi, dovrà avere una specie di bilancia incorporata per stabilire quando il limite viene superato. Un ascensore automatico sicuramente non è cosciente e non stabilisce nulla coscientemente e quindi non ha linee di condotta coscienti. Si può dire, comunque, che aderisce a linee di condotta che dipendono da vari stati che rileva nel mondo e perfino che adatta le linee di condotta a cui aderisce in base ad altri stati di cose che rileva, e così via. Può avere linee di condotta, meta-linee di condotta, meta-meta-linee di condotta, tutte basate su varie e complicate combinazioni di situazioni rilevate – e il tutto senza il minimo accenno di coscienza. Ciò che un ascensore può rilevare e stabilire può essere fatto sicuramente

anche da un cervello umano. Può seguire elaborate linee di condotta inconsce tipo ascensore.

Qual è allora la differenza tra il seguire consciamente o inconsciamente delle linee di condotta? Quando consideriamo le linee di condotta che i nostri corpi seguono inconsciamente, grazie a rilevatori di condizioni «ciechi, meccanici», abbiamo la tentazione di dire che dal *momento che* queste sono linee di condotta inconsce non sono tanto le *nostre* linee di condotta ma quelle dei *nostri* corpi. Le *nostre* linee di condotta sono (per definizione, si potrebbe dire) le nostre linee di condotta coscienti; quelle che formuliamo coscientemente e deliberatamente, con l'opportunità di riflettere (consciamente) sui loro pro e contro, e l'opportunità di adattarle o migliorarle man mano che la situazione si svolge nella nostra esperienza.

Così sembra che quando una condotta viene adottata inizialmente a seguito di una discussione verbale o in risposta ad un'istruzione verbale, essa è *ipso facto* una linea di condotta cosciente, che deve basarsi su un evento coscientemente esperito (Marcel, 1988). Quello che sembra contraddittorio è l'idea che si possa parlare e decidere di seguire una condotta *in*conscia, basandosi su eventi rilevati inconsciamente. Ma c'è, possiamo vedere, una scappatoia: lo status di una tale linea di condotta *potrebbe* cambiare. Con una certa pratica, e qualche dimenticanza strategicamente piazzata, potremmo iniziare da una condotta adottata e seguita consciamente e muoverci gradualmente verso uno stato in cui si segue una condotta *in*conscia rilevando gli aspetti salienti senza esserne coscienti. Questo potrebbe avvenire, ma solo se il legame con la considerazione verbale della linea di condotta venga in qualche modo spezzato.

Questa transizione possibile può essere considerata meglio andando nell'altra direzione. Un soggetto con visione cieca non potrebbe *diventare* cosciente delle esperienze visive nello scotoma, tramite un processo inverso a quello appena immaginato? Dopo tutto, è evidente che nella visione cieca il cervello del soggetto riceve e analizza l'informazione visiva che viene in qualche modo utilizzata nel buon lavoro d'indovinare. Poco dopo che avviene lo stimolo, nel cervello del paziente accade qualcosa che segna l'avvento dello *stato informato*. Se un osservatore esterno (come lo sperimentatore) potesse far in modo di riconoscere tale avvento, potrebbe in linea di principio passare l'informazione al soggetto. Così il soggetto potrebbe arrivare a conoscere «di seconda mano» questi avventi malgrado non sia cosciente di loro «direttamente». E poi il soggetto non dovrebbe

essere in grado, in linea di principio, di «eliminare l'intermediario» e arrivare a riconoscere, proprio come lo sperimentatore, i cambiamenti nelle proprie disposizioni? Al principio, potrà richiedere l'uso di qualche tipo di attrezzatura di auto-osservazione – la stessa attrezzatura su cui si basa lo sperimentatore – ma con il soggetto che ora guarda o ascolta i segnali in uscita.[5]

In altre parole, non sarebbe possibile in linea di principio «chiudere il cerchio di retroazione» e allenare perciò il soggetto a seguire una condotta che tenda a regolare il suo comportamento su cambiamenti di cui non ha avuto un'esperienza («diretta»)? Sollevo la prospettiva di un tale allenamento nel caso della visione cieca come se fosse solo un esperimento ideale, ma in realtà potrebbe essere prontamente trasformato in un esperimento reale. Potremmo provare ad allenare un soggetto con visione cieca a riconoscere quando «indovinare».

I talenti e le disposizioni dei soggetti con visione cieca non sono immutabili; alcuni giorni essi sono in una forma migliore di altri; e migliorano con la pratica, malgrado il fatto che di solito *non* ricevono un immediato feedback dallo sperimentatore su come stanno andando (per eccezioni, vedi Zihl, 1980, 1981). Ci sono svariate ragioni per fare così, principalmente perché tali situazioni sperimentali sono minacciate dalla possibilità di non voluti e non rilevati suggerimenti da parte dello sperimentatore, così le interazioni tra sperimentatore e soggetto sono scrupolosamente minimizzate e controllate. Tuttavia, i soggetti si servono delle imbeccate e dei solleciti che ricevono dagli sperimentatori, e gradualmente si abituano alla pratica altrimenti strana, e non promettente, di fare centinaia o migliaia di congetture su problemi di cui sono convinti di non avere la minima esperienza. (Immagina come ti sentiresti se ti si chiedesse di sedere con un elenco telefonico sulle ginocchia e di indovinare che marca di automobile ha ogni persona nell'elenco, senza nemmeno sentirti dire quando ti capita di rispondere correttamente. Non sembrerebbe un'attività ben motivata per molto tempo, a meno che tu non riceva una qualche credibile assicurazione su come stai andando e perché questa è una bravata che vale la pena di provare.)

Che cosa accadrebbe, allora, se gettassimo al vento tutti gli altri obiettivi scientifici e ci concentrassimo sull'allenamento di qualcuno con la visione cieca, utilizzando tutti i tipi di feedback che potrebbero aiutare per vedere quali risultati si possono ottenere? Supponiamo di iniziare con un ordinario soggetto con visione cieca, che «indovina» tutte le volte che lo imbecchiamo

(la cosiddetta risposta a scelta forzata) e le cui congetture siano migliori di quelle date a caso (se non lo fossero, non sarebbe un soggetto con visione cieca). Il feedback lo porterà presto verso un massimo e, se le congetture si stabilizzassero su un livello di accuratezza piuttosto alto, questo dovrebbe impressionare il soggetto e indurlo a pensare che è in possesso di un talento utile ed affidabile che varrebbe la pena sfruttare. Questo è in effetti lo stato in cui si trovano oggi alcuni soggetti con visione cieca.

Ora supponiamo di iniziare a chiedere al soggetto di andare avanti senza imbeccate – di «indovinare quando indovinare», di indovinare «ogni volta che ne hai voglia» – e di nuovo supponiamo che ogni volta lo sperimentatore fornisca un feedback immediato. Ci sono due risultati possibili:

1. Il soggetto inizia a caso e si ferma lì. Malgrado il fatto che il soggetto sia sicuramente informato dall'avvento dello stimolo, sembra che non ci sia alcun modo tramite il quale il soggetto possa scoprire quando questa informazione sia avvenuta, indipendentemente dalle stampelle di «biofeedback» di cui potremmo fornirlo.
2. Il soggetto alla fine diventa capace di andare avanti senza imbeccate dallo sperimentatore (o da qualunque stampella temporanea di biofeedback), e riesce a fare notevolmente meglio che a caso.

Quale risultato otterremmo in ogni caso particolare è naturalmente una questione empirica e non cerco nemmeno di azzardare un'ipotesi sulla possibilità di ottenere davvero un risultato tipo 2. Forse in ogni singolo caso, il soggetto sarebbe incapace di imparare a «indovinare» correttamente quando indovinare. Ma si noti che, *se* un risultato di tipo 2 dovesse verificarsi, si potrebbe chiedere piuttosto ragionevolmente al soggetto di adottare delle linee di condotta che basano il proprio comportamento su stimoli che egli potrebbe solo indovinare che siano avvenuti. Sia o non sia cosciente di questi stimoli, se l'affidabilità delle sue «congetture» fosse alta, egli potrebbe trattare quegli stimoli alla stregua di ogni esperienza cosciente. Egli potrebbe pensare e decidere su linee di condotta che si basano sul loro avvento come su quello di eventi percepiti coscientemente.

Ma questo lo *renderebbe* in qualche modo cosciente degli stimoli? Come la pensi intuitivamente? Quando ho chiesto a varie persone cosa sarebbero portate a dire, ho ottenuto risposte

contrastanti. La psicologia del senso comune non porta a un chiaro verdetto. Ma un paziente con visione cieca ha parlato per se stesso riferendosi a una circostanza simile. DB, uno dei soggetti studiati da Weiskrantz, ha un'emianopia destra e mostra la classica capacità degli affetti da visione cieca di indovinare accuratamente quando imbeccato. Per esempio, se una luce viene mossa lentamente orizzontalmente o verticalmente attraverso il suo scotoma ed egli viene sollecitato a indovinare «verticale o orizzontale», può ottenere ottimi risultati pur negando di avere la minima coscienza del movimento. Comunque, se la luce viene mossa più rapidamente, il movimento stesso si trasforma in un suggerimento: DB può riferire il movimento volontariamente senza imbeccata e in modo molto accurato, può perfino imitare il movimento con un rapido gesto della mano non appena esso avviene (Weiskrantz, 1988, 1989). Quando gli fu chiesto di commentare questa circostanza, DB disse che *naturalmente* aveva avuto un'esperienza cosciente del movimento – come avrebbe potuto riferirlo sennò? (Anche altri soggetti con visione cieca riferiscono esperienze coscienti di stimoli in rapido movimento.) Possiamo riservarci di pronunciare un giudizio, ma questa risposta non dovrebbe sorprenderci, se l'analisi di Rosenthal del concetto ordinario di coscienza è sui binari giusti. DB non viene solo informato sul movimento della luce, egli *capisce* di essere stato informato; nei termini di Rosenthal, ha un pensiero del secondo ordine sul fatto che ha appena avuto un pensiero del primo ordine.

Otto, il nostro critico, replica:

Ma questo è un altro colpo di mano! Abbiamo sempre saputo che i soggetti con visione cieca sono coscienti del loro indovinare. Tutto questo mostra solo che un soggetto del genere potrebbe sviluppare una notevole abilità a indovinare quando indovinare (e, naturalmente, sarebbe cosciente dell'*azione dell'indovinare*). Giungere a riconoscere che le proprie congetture su quei problemi siano affidabili sarebbe di per sé difficilmente sufficiente per farlo diventare direttamente cosciente dell'evento *su cui* sta indovinando.

Questo suggerisce che per la coscienza visiva sia necessario qualcos'altro. Cosa si potrebbe aggiungere? Certamente la connessione tra la congettura e lo stato *a cui si riferisce*, benché affidabile, sembra piuttosto debole ed effimera. Potrebbe essere

stabilizzata e rafforzata? Quale sarebbe il risultato se i legami di *riferimento* tra la congettura e il suo oggetto fossero moltiplicati?

3. *Nascondi il ditale: un esercizio di potenziamento della coscienza*

Il termine filosofico standard per riferimento è *intenzionalità*, e secondo Elizabeth Anscombe (1965) «deriva da una metafora» dal latino, *intendere arcum in*, che significa *puntare l'arco e la freccia su* (qualcosa). L'immagine del puntare, della direzione è centrale nella maggior parte delle discussioni filosofiche sull'intenzionalità, ma in generale i filosofi hanno barattato il processo complesso di puntamento di una freccia reale con quello di una freccia meramente «logica», una relazione primitiva o fondazionale, resa ancora più misteriosa dalla sua presupposta semplicità. Come *potrebbe* qualcosa nella tua testa puntare questa freccia astratta su una cosa nel mondo?[6] Pensare alla relazione di riferimento come a una relazione logica, astratta può essere in definitiva giusto, ma all'inizio devia l'attenzione dai processi che sono effettivamente implicati nel tenere la mente in un contatto con le cose nel mondo abbastanza stretto, così che su loro si possa *effettivamente* pensare: i processi di badare a, tenersi in contatto con, seguire le tracce e inseguire (Selfridge, inedito). L'effettivo lavorio di puntare qualcosa, «tenerlo focalizzato», implica fare una serie di adattamenti e compensazioni nel corso del tempo, sotto un «feedback di controllo». Ecco perché la presenza di nuvole di distrattori (come il *chaff*, la nube di striscioline metalliche che confonde i sistemi antimissile) può rendere impossibile il puntamento. Focalizzare un bersaglio abbastanza a lungo per identificarlo è un'impresa che richiede più di una singola e momentanea transazione informazionale. Il modo migliore per mantenersi in *contatto* con qualcosa è di tenersi in contatto letterale con essa – afferrarla e non lasciarla andar via, in modo da poterla esaminare finché si vuole con tutto comodo. Il modo migliore immediatamente successivo è di tenersi in contatto figurato, seguendola con gli occhi (e il resto del corpo), non perdendola mai di vista. Questa è un'impresa che può essere realizzata tramite la percezione, naturalmente, ma non solo tramite una percezione passiva; può essere necessario un certo sforzo, una certa pianificazione e, in ogni caso, un'*attività continua* di tenersi in contatto con qualcosa.

Quando ero bambino adoravo giocare a un gioco di gruppo che si chiama: «Nascondi il ditale». Un normalissimo ditale

veniva mostrato a tutti i partecipanti, e poi, mentre tutti gli altri erano fuori della stanza, un bambino «nascondeva» il ditale. Le regole per chi doveva nasconderlo erano chiare: il ditale doveva essere nascosto *in modo perfettamente visibile*. Non doveva essere posto dietro o sotto qualcosa, oppure troppo in alto per essere visto da qualche bambino. In un comune salotto ci sono dozzine di posti in cui si può porre un ditale per farlo apparire mescolato alle cose circostanti come un animale perfettamente mimetizzato. Una volta nascosto, gli altri bambini ritornavano nella stanza e iniziavano la caccia al ditale. Non appena vedevi il ditale dovevi sederti tranquillamente, cercando di non tradire la sua posizione. Di solito succedeva che gli ultimi bambini *fissavano* il ditale varie volte senza in realtà *vederlo*. In quei deliziosi momenti, tutti gli altri potevano vedere che il ditale era proprio di fronte al naso di Betsy, diciamo, ben illuminato e al centro del suo angolo visivo. (In questi momenti a mia madre piaceva dire: «Se fosse un orso, ti morderebbe».) Dalle risatine e dagli urli degli altri bambini, Betsy poteva capire che lo stava fissando – eppure ancora non lo vedeva.

Potremmo esprimerci in questa maniera: anche se qualche stato rappresentazionale nel cervello di Betsy in un certo senso «include» il ditale, nessuno stato percettivo di Betsy si *riferisce* ancora al ditale. Possiamo concedere che uno dei suoi stati coscienti si riferisce al ditale: la sua «immagine di ricerca». Lei può essere tremendamente concentrata nella ricerca del ditale, quello stesso ditale che le è stato concesso di esaminare due o tre minuti fa. Ma nessuna relazione forte di intenzionalità o riferimento si è ancora stabilita fra qualcuno dei suoi stati percettivi e il ditale, anche se in qualcuno degli stati del suo sistema visivo ci potrà essere l'informazione che può rendere possibile a qualcun altro (un osservatore esterno, per esempio, che studia gli stati della sua corteccia visiva) di localizzare o identificare il ditale. Quello che Betsy deve ancora fare è «mirare» al ditale, separarlo come «figura» dallo «sfondo» e identificarlo. Dopo che ciò è avvenuto, Betsy vede realmente il ditale. Il ditale sarà finalmente «nella sua esperienza cosciente» – e ora che è cosciente di esso, potrà finalmente alzare la mano in segno di trionfo – o andare tranquillamente a sedersi con gli altri bambini che hanno già individuato il ditale.[7]

Tali legami finalizzati, guidati da un feedback, corretti in base agli errori, adattati in base ai vantaggi, costituiscono il prerequisito per il tipo di conoscenza che merita questo nome – che può poi servire come un cardine per una linea di condotta,

per esempio. Una volta che ho visto qualcosa in questo senso forte, posso «fare qualcosa che si riferisce ad esso» o fare qualcosa *perché* l'ho visto o *non appena* l'ho visto. Una volta che sono stati identificati, non è difficile normalmente continuare a seguire le tracce dei singoli ditali (a meno che naturalmente non ci si trovi in una stanza piena di ditali durante un terremoto). In circostanze normali, allora, lo status elevato che il ditale raggiunge nel sistema di controllo di Betsy non si mantiene solo per un fugace momento; Betsy manterrà la localizzazione del ditale durante la scoperta o durante il tempo impiegato per riaffermare la sua identità, per controllarlo di nuovo (e di nuovo – se c'è spazio per dubbi). Le cose di cui siamo coscienti in modo più preciso sono quegli oggetti che noi osserviamo apertamente e senza fretta, raccogliendo e integrando i frutti di molti movimenti saccadici, costruendo la conoscenza nel corso del tempo mentre continuiamo a tenere l'oggetto localizzato nello spazio personale. Se l'oggetto svolazza come una farfalla, faremo in modo di immobilizzarlo, «di modo che possiamo osservarlo», e se è ben camuffato nel suo ambiente, dobbiamo fare dei passi – letteralmente, se non dobbiamo toccarlo – per metterlo di fronte ad un sfondo su cui possa risaltare.

Una nostra incapacità a fare questo può impedirci di vedere l'oggetto, in un senso importante e familiare del termine.[8]

I patiti del *birdwatching* spesso tengono un elenco di tutte le specie di uccelli che hanno visto. Supponiamo che tu ed io fossimo tra loro; quando entrambi sentiamo gli uccelli cantare tra gli alberi sopra di noi, io alzo la testa e dico: «Io lo vedo – e tu?». Tu guardi esattamente dove io sto guardando, eppure dici, sinceramente: «No, non lo vedo». Io scriverò questo uccello nel mio elenco; tu non lo farai malgrado tu possa essere moralmente certo che la sua immagine sia oscillata varie volte nelle tue fovee.

Cosa diresti? Il ditale era in un certo senso «presente» nella coscienza di Betsy prima che lo scovasse? L'uccello era presente nello «sfondo» della tua coscienza o non era affatto presente? Portare qualcosa nel primo piano della coscienza equivale a metterlo in una posizione in cui possa essere riferito, ma cosa è necessario per portare qualcosa nello sfondo dell'esperienza cosciente (e non meramente nello sfondo dell'ambiente visibile)? Il ditale e l'uccello erano indubbiamente presenti nell'ambiente visibile – ma non è questo il punto. Presumibilmente non è sufficiente che la luce riflessa dall'oggetto entri meramente negli occhi, ma quale effetto ulteriore deve avere la luce

riflessa – di che altro si deve accorgere il cervello – affinché l'oggetto passi dal rango delle cose a cui si risponde in modo semplicemente inconscio a quelle presenti nello sfondo dell'esperienza cosciente?

Il modo di rispondere a questi intoppi «del punto di vista in prima persona» è di ignorare il punto di vista in prima persona e esaminare quello che può essere appreso dal punto di vista in terza persona. Nei capitoli 8-10, abbiamo esaminato un modello di produzione di atti linguistici che dipendeva da un processo Pandemonio in cui il definitivo accoppiamento del contenuto con l'espressione era il risultato di varie competizioni, la costruzione, lo smantellamento e la ricostruzione di coalizioni. I contenuti che sono entrati nella rissa ma non sono riusciti a rimanervi a lungo potrebbero inviare una specie di effetto a colpo unico «balistico» che si propaga nel sistema, ma sarebbero quasi *irriferibili*. Se un evento non si protrae, qualunque tentativo di riferirlo, se avviato, o sarà abortito o sfuggirà al controllo, non avendo nulla su cui correggere se stesso. Affinché ci sia la riferibilità, è necessario che ci sia la capacità di identificare e reidentificare l'effetto. Possiamo seguire lo sviluppo della riferibilità in molte varietà di allenamento, che in qualche modo ci ricordano l'allenamento a cui abbiamo immaginato di sottoporre il nostro paziente con visione cieca: i risultati dell'allenamento del palato negli assaggiatori di vino, l'allenamento dell'orecchio nei musicisti, e simili – o il semplice esperimento con le corde di chitarra descritto nel capitolo 3.

Si considerino, per esempio, le istruzioni che vengono date agli accordatori di pianoforte apprendisti. Viene loro detto di ascoltare i «battimenti» mentre schiacciano il tasto che stanno accordando assieme a un tasto di riferimento. *Quali battimenti*? All'inizio, molti apprendisti non sono capaci di discernere nella loro esperienza auditiva nulla che corrisponda a quei battimenti – ciò che sentono è qualcosa che descriverebbero come una specie di brutto suono non strutturato o stonato. Alla fine, però, se l'allenamento ha successo, arrivano a essere in grado di isolare, nella loro esperienza auditiva, l'interferenza dei battimenti e a notare come la sequenza dei battimenti si modifica in seguito alla loro azione di avvitare il pirolo con la chiave di accordatura. Riescono così ad accordare facilmente il piano accordando i battimenti. Quello che di solito dicono – e tutti noi possiamo confermarlo con episodi simili nella nostra esperienza – è che a seguito dell'allenamento *la loro esperienza cosciente si è modificata*. Più specificatamente, è stata accresciuta: ora sono coscienti di cose di cui prima non erano coscienti.

In un certo senso è vero che essi hanno ascoltato sempre i battimenti. È la loro interferenza, dopo tutto, che compone la stonatura di cui erano sicuramente coscienti. Ma precedentemente erano incapaci di rilevare queste componenti nella loro esperienza; per questo si potrebbe dire che questi fattori hanno *contribuito* ma non erano essi stessi *presenti* nell'esperienza. Lo *status* funzionale di tali contributi prima dell'allenamento era lo stesso di quello degli eventi che avvengono nella visione cieca: il soggetto è incapace di riferire i contributi particolari o di basare una linea di condotta sul loro avvento, ma i risultati di questi contributi possono tuttavia essere evidenziati nel comportamento del soggetto, per esempio nella capacità del soggetto di rispondere ad appropriate domande. Quello che sto suggerendo è che non ci deve essere nulla di più di questo nello sfondo dell'esperienza. Ora non è impossibile, come abbiamo visto, che un legame rafforzato del tipo che abbiamo appena descritto per gli accordatori di pianoforte e gli assaggiatori di vino possa essere costruito in un soggetto con visione cieca fino al punto in cui dichiari, e sia prontamente creduto, che è *diventato* cosciente degli stimoli – perfino nel primo piano della sua coscienza – che prima poteva solo indovinare.

Non corriamo troppo [dice Otto]. Ecco un'altra obiezione: tu immagini che il soggetto con visione cieca possa imparare ad usare la sua visione cieca in queste nuove maniere e che forse ciò potrebbe dargli una *sorta* di coscienza degli eventi che si verificano nel suo campo cieco, ma questo lascia ancora qualcosa fuori. La coscienza non sarebbe coscienza *visiva*; non sarebbe come il *vedere*. Le «qualità fenomeniche» o *qualia* della visione cosciente mancherebbero, anche se il soggetto con visione cieca riuscisse a fare tutte queste mosse funzionali.

Forse sì e forse no. Ma cosa sono le «qualità fenomeniche» o *qualia*? All'inizio sembrano incredibilmente ovvie – è il modo in cui l'aspetto, l'odore, il suono delle cose sembrano a noi – ma hanno una tendenza a modificare il loro *status* o a svanire se esaminate. Nel prossimo capitolo seguiremo questi sospetti fin nel folto delle discussioni filosofiche, ma prima dovremmo dare un'occhiata migliore ad alcune proprietà che *non* sono qualità fenomeniche, ma potrebbero essere scambiate per esse.

4. *Visione protesica: indipendentemente dall'informazione, che cosa manca ancora?*

Il soggetto di Weiskrantz, DB, *vede* il movimento? Be', sicuramente non lo *ode* né lo *sente*. Ma abbiamo a che fare con la vista? Ha le «qualità fenomeniche» della vista? Weiskrantz dice:

> Quando la «cospicuità» dello stimolo aumenta, il paziente può insistere nel dire che ancora non «vede», ma che ora ha l'«impressione» che qualcosa sia lì. In alcuni casi, se lo stimolo è ancora più cospicuo, si può raggiungere un punto in cui il soggetto dice di «vedere», ma l'esperienza non è veridica. Per esempio, DB «vede» in risposta a uno stimolo in rapido movimento, ma non vede un oggetto coerente che si muove, riferisce piuttosto complesse configurazioni di «onde». Altri soggetti riferiscono «ombre scure» che emergono man mano che la luminosità e il contrasto raggiungono alti livelli. (1988, p. 189)

L'oggetto che si muove rapidamente non è percepito da DB con un colore o una forma, e allora? Come abbiamo dimostrato con noi stessi nel capitolo 2, nell'esperimento con la carta da gioco tenuta nella visione periferica, possiamo sicuramente vedere la carta senza essere capaci di identificare né il suo colore né il suo seme. E questa è visione normale, non visione cieca, quindi dovremmo stare attenti a negare al soggetto l'esperienza visiva solo su queste basi.

Possiamo porre in modo più vivido la questione se questo modo anormale di ottenere informazioni su oggetti visibili sia o no una varietà di *esperienza visiva*, allontanandoci in modo più radicale dalla vista normale. Vari congegni protesici sono stati inventati per dare la «vista» ai ciechi, e alcuni di loro fanno sorgere proprio le tematiche che stiamo discutendo. Quasi venti anni fa, Paul Bach-y-Rita (1972) ha sviluppato vari congegni con piccole telecamere a bassa risoluzione che potevano essere sistemate su montature di occhiali. I segnali di queste telecamere a bassa risoluzione, una matrice 16 per 16 o 20 per 20 di pixel in «bianco e nero», venivano diffusi sulla schiena o sulla pancia dei soggetti da una griglia di punture meccaniche o elettriche chiamate «tattori».

Dopo solo poche ore di allenamento, i soggetti ciechi che utilizzavano questo congegno potevano imparare ad interpretare le configurazioni delle punture sulla loro pelle, un po' come noi possiamo interpretare le lettere tracciate sulla nostra schiena dal dito di un altro. La risoluzione è bassa, ma anche

11.3

Un soggetto cieco con un sistema elettrico portatile a 16 linee. La telecamera è sistemata sulla montatura degli occhiali; un piccolo fascio di fili la connette a un circuito che produce uno stimolo elettrico (nella mano destra). La matrice di 256 elettrodi d'argento concentrici viene tenuta nella mano sinistra.

così i soggetti potevano imparare a leggere dei segni e identificare oggetti e perfino volti umani, come possiamo vedere *guardando* questa fotografia del segnale così come appare sul monitor di un oscilloscopio.

Ne risultava certamente un'esperienza percettiva cosciente prodotta da una protesi, ma dal momento che i segnali in ingresso arrivavano alla pancia o alla schiena dei soggetti e non alle loro retine, si trattava di *visione*? Aveva le «qualità

11.4

Così appare sull'oscilloscopio una rappresentazione a 400 pixel del volto di una donna. I soggetti possono identificare correttamente configurazioni di stimoli di questo livello di complessità.

fenomeniche» della visione o solo della sensazione tattile?

Ricordati di uno degli esperimenti del capitolo 3; è piuttosto facile che il tuo punto di vista tattile si sposti nella punta della matita, permettendoti di sentire la tessitura con la punta della matita, mentre ovviamente rispondi alle vibrazioni della matita sulle dita. Così non dovrebbe sorprenderci di venir a sapere che un effetto simile, sebbene più estremo, veniva provato dai soggetti di Bach-y-Rita. Dopo un breve periodo di allenamento, la loro consapevolezza delle punture sulla schiena scompariva; il cuscinetto di pixel diventava trasparente, si potrebbe dire, e il

punto di vista dei soggetti si spostava nel punto di vista della telecamera, montata sulla loro testa. Una dimostrazione sorprendente di quanto fosse intenso questo spostamento di prospettiva è data dal comportamento di un soggetto, a lungo allenato, che aveva una telecamera dotata di zoom azionabile da un pulsante (pp. 98-99). La matrice di punture veniva applicata sulla sua schiena e la telecamera era montata sulla sua testa. Quando lo sperimentatore senza avvertirlo azionava lo zoom, facendo ingrandire o «avvicinare» improvvisamente l'immagine sulla sua *schiena*, il soggetto istintivamente si tirava *indietro, alzando le braccia a protezione della testa*. Un'altra dimostrazione sorprendente della trasparenza delle punture sta nel fatto che i soggetti che sono stati allenati con una griglia di punture sulla schiena potevano adattarsi quasi immediatamente allo spostamento della griglia di punture sulla loro pancia (p. 33). Eppure, come Bach-y-Rita nota, seguitavano a grattarsi se avevano un prurito sulla schiena – non dicevano di «vederlo» – ed erano perfettamente in grado di considerare le punture come punture, se veniva loro richiesto.

Queste osservazioni sono stimolanti ma inconcludenti. Si potrebbe sostenere che quando l'uso dei segnali in ingresso forniti dal congegno sia divenuto una seconda natura i soggetti vedano realmente oppure, al contrario, che solo alcune delle più importanti caratteristiche «funzionali» della vista sono riprodotte dalla protesi. E le altre «qualità fenomeniche» della visione? Bach-y-Rita racconta di aver mostrato, per la prima volta nella loro vita, a due soggetti allenati, studenti universitari maschi ciechi, delle fotografie di donne nude dalla rivista «Playboy». Essi erano contrariati dal fatto che «l'esperienza non aveva componenti emotive, nessun sentimento piacevole veniva risvegliato, malgrado entrambi potessero descrivere abbastanza bene il contenuto delle fotografie. Ciò disturbava profondamente i due giovanotti, che erano consapevoli del fatto che simili fotografie contenevano una componente emotiva per i loro amici che vedevano normalmente» (p. 145).

Così le protesi di Bach-y-Rita non producono *tutti* gli effetti della vista normale. Alcune deficienze possono essere dovute alla sorprendente differenza di velocità del flusso dell'informazione. La vista normale ci informa sulle proprietà spaziali degli oggetti nel nostro ambiente a grande velocità e quasi con qualsiasi grado di dettaglio desiderato. Non sorprende che le informazioni spaziali a bassa risoluzione inviate al cervello attraverso un'interfaccia con la pelle non riescano a sollecitare

quelle reazioni che vengono provocate nelle persone normalmente vedenti quando i loro sistemi visivi vengono inondati da segnali in ingresso.[9] Quanto piacere ci aspetteremmo che un uomo con vista normale riceverebbe *guardando* traduzioni a bassa risoluzione – dai un'occhiata alla figura 11.4 – di fotografie di una bella donna?

Non è chiaro quali cambiamenti ci sarebbero se qualcuno riuscisse a migliorare la «velocità in baud»[10] della visione protesica per arrivare a valori simili a quella normale. Forse basterebbe incrementare semplicemente la quantità e la velocità dell'informazione, fornendo in un certo senso al cervello una mappa di bit più accurata, per produrre le delizie che mancano. O piuttosto alcune di esse. I nati ciechi si troverebbero in una condizione tremendamente più svantaggiata rispetto alle persone che hanno perso recentemente la vista, poiché non hanno dentro di sé nessuna delle associazioni tipicamente visive che senza dubbio giocano un ruolo importante nel deliziare l'esperienza dei normali vedenti, *ricordando* loro precedenti esperienze visive. Potrebbe anche essere che una parte del piacere di cui godiamo nelle esperienze visive sia un prodotto collaterale di antiche tracce fossili di una precedente economia nei nostri sistemi nervosi, un punto sollevato nel capitolo 7 che sarà indagato ulteriormente nel prossimo capitolo.

Le stesse considerazioni si applicano alla visione cieca e a qualunque miglioramento si voglia immaginare nelle abilità dei soggetti con visione cieca. Le discussioni sulla visione cieca hanno avuto la tendenza a ignorare la povertà delle informazioni che i soggetti racimolano dai loro campi ciechi. Una cosa è essere in grado di indovinare, su sollecitazione, se un quadrato o un cerchio è stato appena mostrato nel campo cieco; un'altra, essere in grado di indovinare in dettaglio, su sollecitazione, che cosa è appena avvenuto fuori della finestra.

Possiamo utilizzare ciò che abbiamo appreso sulla visione protesica per guidare la nostra immaginazione su cosa proverebbe un soggetto con visione cieca se dovesse riappropriarsi di molte delle *funzioni* della vista. Proviamo a immaginare di incontrare una persona corticalmente cieca, che, dopo assiduo allenamento, (1) abbia trasformato la sua capacità di indovinare quando indovinare in una seconda natura, (2) possa giocare a: «Nascondi il ditale» con gli altri virtuosi della visione cieca, e (3) sia in qualche modo riuscito a migliorare la velocità e i dettagli delle sue congetture di qualche ordine di grandezza. Lo incontriamo mentre legge il giornale e ridacchia sui fumetti,

e gli chiediamo delle spiegazioni. Ecco tre scenari, in ordine ascendente di plausibilità:

1. «Sto solo indovinando, ovviamente! Non posso vedere un cavolo, lo sai bene, ma ho imparato a indovinare quando indovinare, e ora, per esempio, immagino che mi stai rivolgendo un gesto maleducato e stai storcendo la faccia per esprimere la tua totale perplessità.»
2. «Ciò che è iniziato come semplici congetture si è trasformato gradualmente in qualcos'altro e a cui ora mi affido di più. Li potremmo chiamare *presentimenti*. Mi sono trovato in una situazione in cui improvvisamente *sapevo* che qualcosa entrava nel mio campo visivo cieco. Potevo allora esprimere la mia conoscenza e agire di conseguenza. Inoltre, avevo poi una meta-conoscenza sul fatto che ero capace di tali presentimenti e potevo usare questa meta-conoscenza nella pianificazione delle mie azioni e nella definizione delle mie linee di condotta. Ciò che è iniziato come congetture coscienti si è trasformato in presentimenti coscienti, ed ora arrivano con una velocità e una forza tale che non posso neppure separarli! Ma ancora non posso vedere un cavolo! Non come ero abituato a fare! No, non è affatto come vedere.»
3. «In realtà, assomiglia *molto* al vedere. Ora agisco senza sforzo nel mondo sulla base delle informazioni che racimolo dall'ambiente circostante tramite i miei occhi. O, se voglio, posso essere cosciente su ciò che ottengo tramite gli occhi. Senza la minima esitazione reagisco ai colori delle cose, alle loro forme e collocazioni, e non ho più quel senso di sforzo che provavo quando stavo sviluppando questi talenti, ora mi sembrano una seconda natura.»

Eppure possiamo ancora immaginare che il nostro soggetto dica che manca qualcosa:

«I *qualia*. I miei stati percettivi hanno qualia, naturalmente, perché sono degli stati coscienti, ma prima della perdita della vista avevano qualia *visivi* e ora non più, malgrado tutto il mio allenamento.»

Ti potrà sembrare che quest'affermazione sia sensata, che è proprio quello che ti aspettavi che il soggetto dicesse. Se è così, il resto del capitolo è per te: un esercizio per scuotere questa convinzione. Se stai già incominciando a dubitare che questo discorso sui

qualia abbia un qualche senso, hai probabilmente anticipato alcune delle svolte che la nostra storia sta per prendere.

5. *«Riempire» contro scoprire*

> Ma, per il fatto che esiste questo sentimento di estraneità, non si può dire: ogni oggetto, che ben conosciamo e che ora non ci appare estraneo, ci dà un senso di familiarità. – Pensiamo, per così dire, che il posto prima riempito dal sentimento di estraneità debba essere occupato in *qualche modo*.
>
> LUDWIG WITTGENSTEIN (1953, I, 596)

Nel capitolo 2, abbiamo visto che una delle ragioni per credere al dualismo era la sua promessa di fornirci il «materiale di cui son fatti i sogni» – la mucca viola e altri figmenti della nostra immaginazione. E nel capitolo 5 abbiamo visto le confusioni che sorgono dalla presupposizione naturale ma errata che *dopo* essere arrivato a una discriminazione o a un giudizio, il cervello *ri-presenti* il materiale su cui è basato il giudizio, per il piacere del pubblico nel Teatro Cartesiano, inserendo i colori. Possiamo rintracciare quest'idea dell'*inserimento* o *riempimento* anche nel pensiero dei più sofisticati teorici, ed essa tradisce quasi sempre un incompleto superamento del materialismo cartesiano. Il divertente è che quelli che usano questi termini quasi sempre sono tutt'altro che sprovveduti, ma, poiché trovano i termini irresistibili, pensano di mettersi al riparo ricorrendo alle virgolette.

Per esempio, quasi tutti parlano del cervello che «riempie» il punto cieco (in tutti gli esempi seguenti il corsivo è mio):

> [...] il ben conosciuto fenomeno neurologico del «*riempimento*» soggettivo della porzione mancante dell'area cieca nel campo visivo. (Libet, 1985b, p. 567.)

> [...] puoi localizzare il tuo punto cieco, e dimostrare anche come una configurazione viene «*riempita*» o «completata» nel punto cieco [...]. (Hundert, 1987, p. 427.)

C'è anche un «riempimento» auditivo. Quando ascoltiamo un discorso, possiamo «riempire» le lacune nei segnali acustici – per esempio nell'«effetto ristorativo dei fonemi» (Warren, 1970). Ray Jackendoff si esprime così:

Si consideri, per esempio, la percezione di un discorso con segnali in ingresso difettosi o rumorosi – diciamo, per la presenza di un aereo a reazione in funzione o per un collegamento telefonico difettoso [...] Ciò che si ricostruisce [...] non è solo un significato inteso ma anche una struttura fonologica: si «ascolta» più di quello che il segnale effettivamente contiene [...] In altre parole, l'informazione fonetica viene «*riempita*» dalle strutture di livello superiore così come dai segnali acustici; e malgrado ci sia una differenza nel modo in cui viene derivata, non c'è nessuna differenza qualitativa nella stessa struttura completata. (Jackendoff, 1987, p. 150)

E quando leggiamo un testo, avviene qualcosa di simile (benché visivo). Come si esprime Bernard Baars:

Troviamo fenomeni simili nel ben conosciuto «effetto del correttore di bozze», la ben nota difficoltà a rilevare gli errori di stampa nelle bozze tipografiche perché la mente «inserisce» l'informazione corretta. (Baars, 1988, p. 173)

Howard Margolis aggiunge un'annotazione non contrastata sull'intera faccenda dell'«inserimento»:

I dettagli «inseriti» sono normalmente corretti. (Margolis, 1987, p. 41.)

Un riconoscimento tacito del sospetto che ci sia qualcosa che puzza nell'idea del riempimento viene manifestato chiaramente in questa descrizione del punto cieco fatta dal filosofo C.L. Hardin, nel suo libro *Color for Philosophers*:

Copre un'area con un diametro di 6 gradi visivi, sufficiente a contenere le immagini di dieci lune piene poste l'una accanto all'altra, eppure non c'è nessun buco nella regione corrispondente del campo visivo. Questo perché l'occhio del cervello lo *riempie* con qualsiasi cosa vede nelle regioni circostanti. Se il contorno è azzurro lo *riempie* d'azzurro; se è a scacchi, noi non siamo consapevoli di alcuna discontinuità nell'estensione della scacchiera. (1988, p. 22.)

Hardin non riesce proprio a dire che il cervello riempie il punto cieco con un disegno a scacchi, poiché questo suggerisce, sicuramente, un lavorio di «ricostruzione» piuttosto sofisticato, come l'elegante «rammendo invisibile» per cui sei ben disposto a sborsare una cifra notevole quando devi riempire un buco nella tua giacca a spina di pesce: tutte le linee e tutte le tonalità

di colore combaciano perfettamente nel confine tra il vecchio e il nuovo. Sembra che riempire con l'azzurro sia una cosa – tutto quello che servirebbe è uno o due colpi di pennello cerebrale intinto nel colore giusto; ma riempire a scacchi è un'altra cosa, ed è più di quanto sia disposto a dire.

Ma, come il commento di Hardin ci ricorda, noi non siamo consapevoli del nostro punto cieco sia quando guardiamo una superficie a scacchi sia quando guardiamo una superficie a tinta unita, così qualunque cosa serva a produrre questa inconsapevolezza essa può essere facilmente realizzata dal cervello in entrambi i casi. «Noi non siamo consapevoli di alcuna discontinuità», come egli dice. Ma se il cervello non deve riempire la lacuna con degli scacchi, perché dovrebbe preoccuparsi di riempirla con l'azzurro?

In nessun caso, presumibilmente, il «riempimento» è una questione di riempimento letterale – del tipo che richiederebbe qualcosa di simile ad un pennello. (Questa era la morale della storia del CADBLIND Mark II nel capitolo 10.) Sono sicuro che nessuno pensi che il «riempimento» sia qualcosa che il cervello realizzi coprendo realmente una superficie spaziale con un *pigmento*. Sappiamo che l'immagine reale e capovolta sulla retina è l'ultimo stadio della visione in cui ci sia qualcosa di colorato nel modo non problematico in cui un'immagine di un film è colorata. Poiché non c'è un occhio della mente letterale, non c'è bisogno di pigmento nel cervello.

Tanto peggio per il pigmento, allora. Tuttavia, potremmo essere inclini a pensare che nel cervello avvenga qualcosa che sia *in qualche modo* strettamente analogo al coprire un'area con un pigmento – altrimenti non parleremmo affatto di «riempire». È questo avvenimento speciale, qualunque esso sia, che si verifica, evidentemente, nel «mezzo» speciale dell'esperienza visiva o auditiva. Come Jackendoff dice, parlando del caso auditivo, «si "ascolta" più di quello che il segnale effettivamente contiene» – ma si noti che egli mette «ascolta» tra virgolette. Che cosa potrebbe essere ciò che è *presente* quando si «ascoltano» suoni che riempiono i momenti di silenzio o si «vedono» colori che si estendono su superfici vuote? Sembrerebbe che in questi casi ci sia qualcosa che è presente, qualcosa che deve essere fornito dal cervello (tramite il «riempimento»). Come dovremo chiamare quest'entità sconosciuta? Chiamiamola *figmento*. La tentazione, allora, è di supporre che ci sia qualcosa, fatta di questo figmento, che è presente quando il cervello «riempie» e che manca quando non si deve preoccupare di

«riempire». In questa forma così debole, l'idea del figmento non è molto attraente (almeno penso che non dovrebbe esserlo). Non siamo così sprovveduti: non esiste una cosa come il figmento. Il cervello non produce figmento; il cervello non usa figmento per riempire le lacune; il figmento è solo un figmento della mia immaginazione. Tanto peggio per il figmento, allora! Ma allora, cosa significa «riempire», cosa *potrebbe* significare, se non significa riempire con un figmento? Se non c'è un mezzo come il figmento, che differenza c'è tra il «riempire» e il non preoccuparsi di riempire?

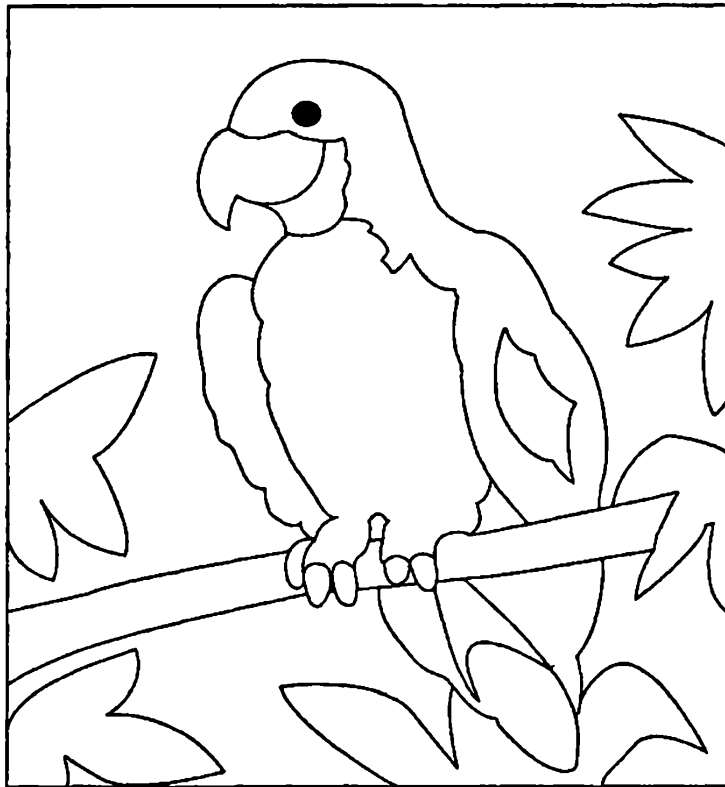
Nel capitolo 10 abbiamo visto come un sistema CAD possa rappresentare i colori associando un numero identificativo del colore a ogni pixel o a ogni regione delimitata dell'oggetto raffigurato, e abbiamo visto anche come il CADBLIND Mark II possa ricercare, o rilevare, i colori leggendo tale codice. Questo processo ci ricorda un passatempo dei bambini, «colora seguendo i numeri», che ci fornisce un semplice analogo dei processi di rappresentazione che devono avvenire, o potrebbero avvenire, nel cervello. La figura 11.5 è una rappresentazione con informazioni sulle forme, ma affatto priva di informazioni sui colori.

Paragoniamola alla figura 11.6, che contiene informazioni sui colori nella forma di un codice numerato. Se prendessimo

11.5

11.6

1. Blu
2. Verde
3. Arancione
4. Rosso
5. Viola
6. Giallo

11.7

delle matite colorate e seguissimo le indicazioni su come riempire gli spazi con i colori, potremmo trasformare la figura 11.6 in un'altra specie di rappresentazione «riempita» – una in cui le regioni sarebbero riempite di colore reale, pigmento reale.

C'è ancora un altro modo in cui il colore si potrebbe «inserire», pixel per pixel, tramite una mappa codificata di bit, come nella figura 11.7.

Entrambe le figure 11.6 e 11.7 sono rappresentazioni «riempite» (a differenza della figura 11.5, per esempio), poiché qualunque procedura che deve essere informata sul colore di una regione può, tramite un'ispezione meccanica di quella regione, estrarre quell'informazione. Questo è un riempimento puramente informazionale. I sistemi sono, naturalmente, interamente arbitrari. Noi possiamo costruire facilmente un numero indefinito di sistemi di rappresentazione funzionalmente equivalenti – che implicano differenti sistemi di codificazione, o mezzi differenti.

Se vuoi fare un disegno colorato sul tuo personal computer usando il programma *PC-Paintbrush*, lo schermo viene rappresentato nella macchina come una mappa di bit nella «memoria di transito», in modo analogo alla figura 11.7, ma quando salvi il tuo disegno su disco, un algoritmo di compressione lo traduce in qualcosa che è più simile alla figura 11.6. Divide l'area in regioni colorate nello stesso modo e salva i confini delle regioni e i rispettivi numeri del colore in un file d'archivio.[11] Il file d'archivio è una rappresentazione altrettanto accurata della mappa di bit, ma, grazie alla generalizzazione regionale e all'etichettazione non ripetuta di ogni regione, è un sistema molto più efficiente.

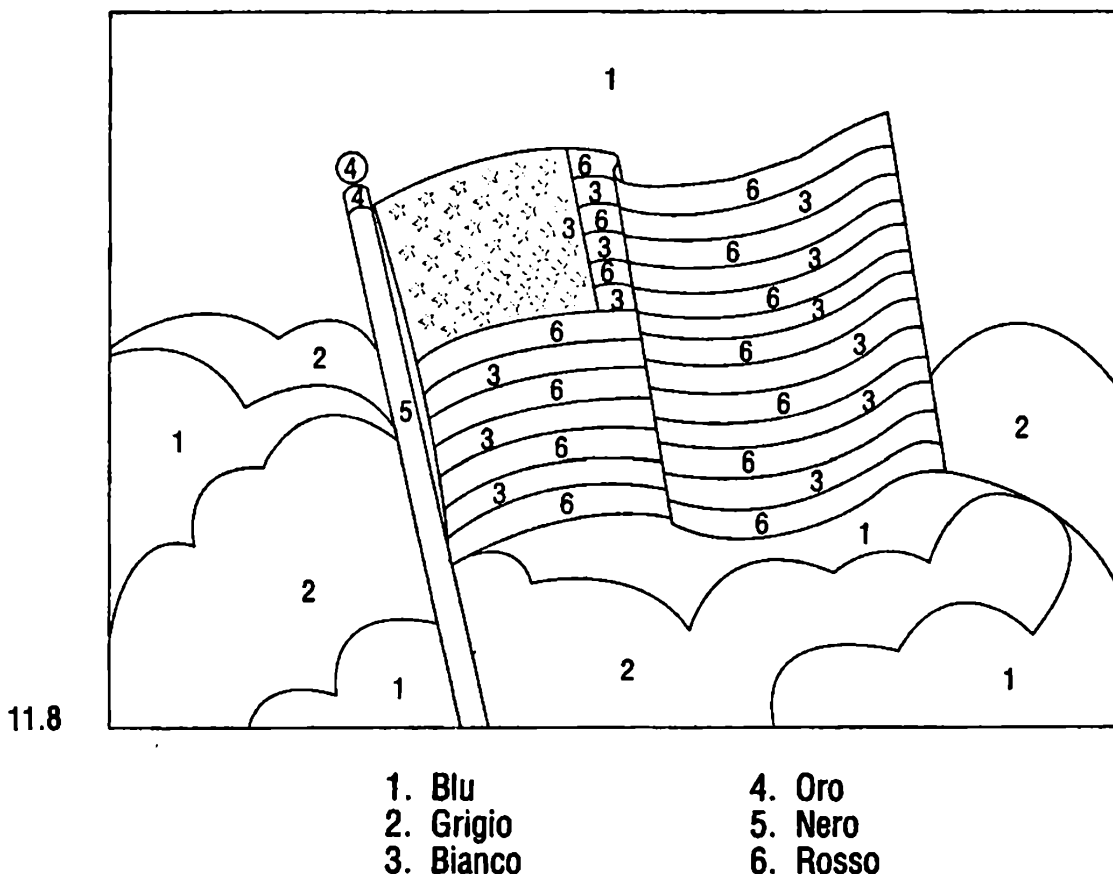
Una mappa di bit, etichettando esplicitamente ogni pixel, è una forma di ciò che potremmo chiamare una rappresentazione *approssimativamente continua* – l'approssimazione è in funzione della dimensione dei pixel. Una mappa di bit non è letteralmente un'immagine, ma solo una matrice di valori, una sorta di ricetta per formare un'immagine. La matrice può essere immagazzinata in qualunque sistema che conserva l'informazione sulla collocazione. Il videotape è un altro mezzo di rappresentazione approssimativamente continua, ma ciò che immagazzina sul nastro non sono letteralmente immagini, ma ricette (ad un livello granulare differente) per formare immagini.

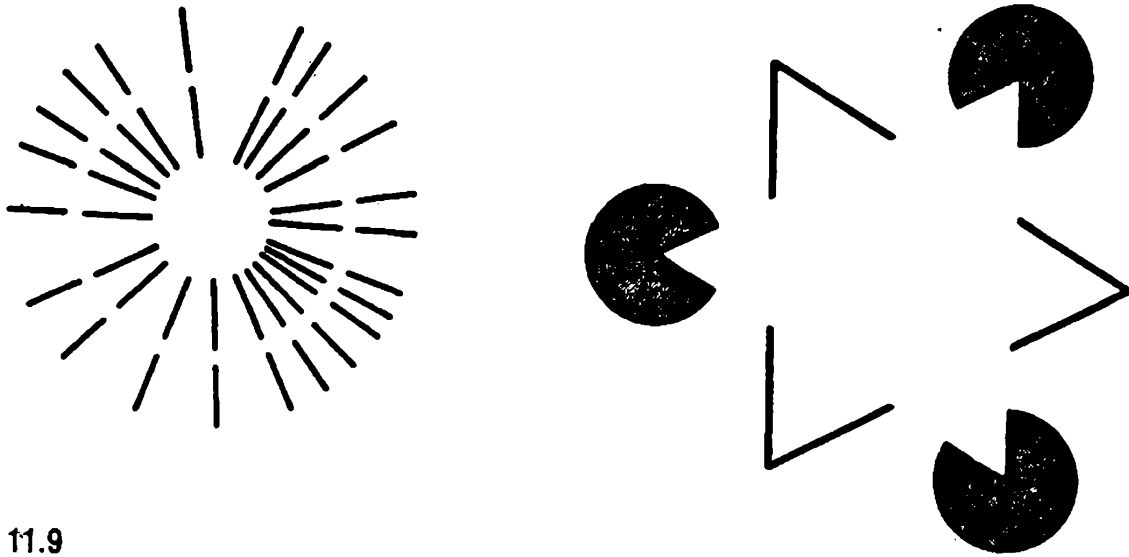
Un altro modo per salvare l'immagine sullo schermo del calcolatore sarebbe quello di farne una fotografia a colori e conservare l'immagine, diciamo, in una diapositiva formato

35 mm; questo sistema differisce dagli altri per un aspetto importante, e ovvio; c'è una tinta reale che riempie letteralmente una regione di spazio reale. Come la mappa di bit, questa è una rappresentazione approssimativamente continua delle regioni spaziali raffigurate (continua fino alla grana del film – su scala sufficientemente piccola, diventa simile a un insieme di pixel o granulare). Ma a differenza della mappa di bit, il colore è usato per rappresentare il colore. Anche un negativo a colori usa il colore per rappresentare il colore ma in modo invertito.

Ecco, allora, tre modi differenti per inserire l'informazione cromatica: colore tramite numeri, come nella figura 11.6 o un file d'archivio, colore tramite una mappa di bit, come nella figura 11.7 o la memoria di transito o il videotape, e colore tramite colori. Il colore tramite numeri è, sotto un certo aspetto, un modo per inserire l'informazione cromatica ma raggiunge la sua efficienza, in paragone agli altri, proprio perché *non* si preoccupa di inserire esplicitamente i valori per ogni pixel. Ora in quale di questi (o altri) sensi il cervello «riempie» il punto cieco? Nessuno pensa che il cervello usi numeri in registri per codificare i colori, ma questa è una digressione irrilevante. Possiamo pensare che i numeri nei registri stiano per qualunque sistema di grandezza, qualunque sistema di «vettori» che un cervello potrebbe impiegare come un «codice» per i colori; potrebbe trattarsi di frequenze di scariche neurali o qualche sistema di indirizzi o collocazioni in reti neurali o qualunque altro sistema di variazioni fisiche nel cervello. I numeri nei registri hanno la simpatica proprietà di conservare le relazioni tra grandezze fisiche pur rimanendo neutrali su qualunque proprietà «intrinseca» di tali grandezze, così possono stare per qualsiasi grandezza fisica nel cervello che «codifica» i colori. Benché i numeri possano essere usati in modi completamente arbitrari, possono essere usati anche in modi non arbitrari, per riflettere le relazioni strutturali tra colori che sono state scoperte. Il noto «solido dei colori» in cui la tonalità, l'intensità e la luminosità sono le tre dimensioni lungo le quali variano i *nostri* colori,[12] è uno spazio logico idealmente adatto a un trattamento numerico – ogni trattamento numerico che riflette le relazioni di reciprocità, di opposizione e di complementarità, eccetera, che la vista umana effettivamente mostra. Più apprendiamo su come il cervello codifica i colori, più potente e meno arbitrario sarà un modello numerico della visione cromatica umana che potremo escogitare.

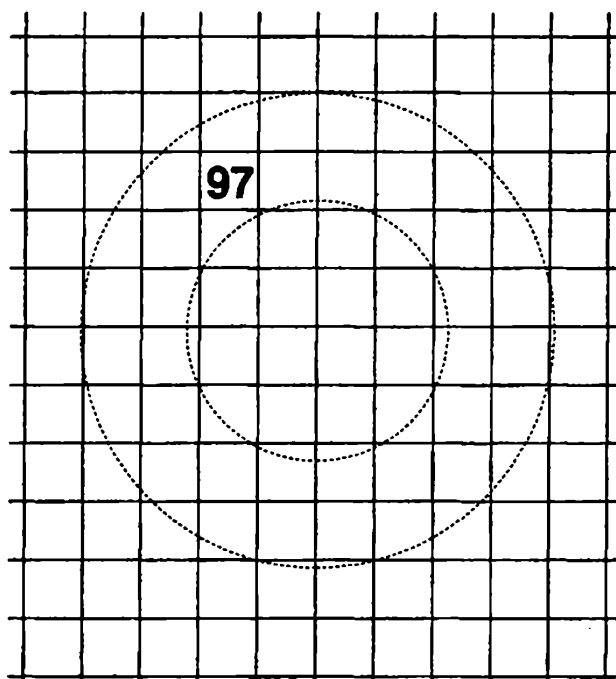
Il problema che si incontra parlando della «codificazione» cerebrale dei colori tramite intensità o grandezze di un tipo o dell'altro è che può indurre l'incauto a pensare che questa codificazione debba alla fine essere decodificata, «riportandoci al colore». Questa è una strada – molto frequentata – che ci riporta al figmento: si immagina che il cervello possa immagazzinare inconsciamente le sue informazioni enciclopediche sul colore in un formato simile a quello della figura 11.8 ma che poi si organizzi per «decodificare» la rappresentazione in «veri colori» in occasioni speciali – come quando si mette in funzione una videocassetta per proiettare i colori reali sullo schermo. Esiste certamente una differenza fenomenologica tra il semplice ricordare la frase che ci informa che la bandiera americana è rossa, bianca e blu, e l'immaginarla effettivamente «a colori», «vedere» (con l'occhio della mente) che è rossa, bianca e blu. Se questa differenza fenomenologica ispira alcune persone a postu-

11.8

lare il figmento, un caso ancor più stringente ci viene offerto dal fenomeno della diffusione al neon del colore (van Tuijl, 1975), di cui possiamo vedere un esempio nel retro della copertina di questo libro.

Ti sembra che l'anello definito dalle linee rosse si riempia di rosa, ma non c'è nulla che sia rosa sulla pagina, né inchiostro mal distribuito né strane riflessioni di luce. In altre parole, sulla tua retina non c'è il rosa, ma solo righe rosse e nere. Ora come si potrebbe spiegare questa illusione? Un circuito cerebrale, specializzato nelle forme, viene indotto erroneamente a distinguere una particolare regione delimitata: l'anello con i suoi «contorni soggettivi». Molte altre figure, come le seguenti, producono contorni soggettivi.

11.9

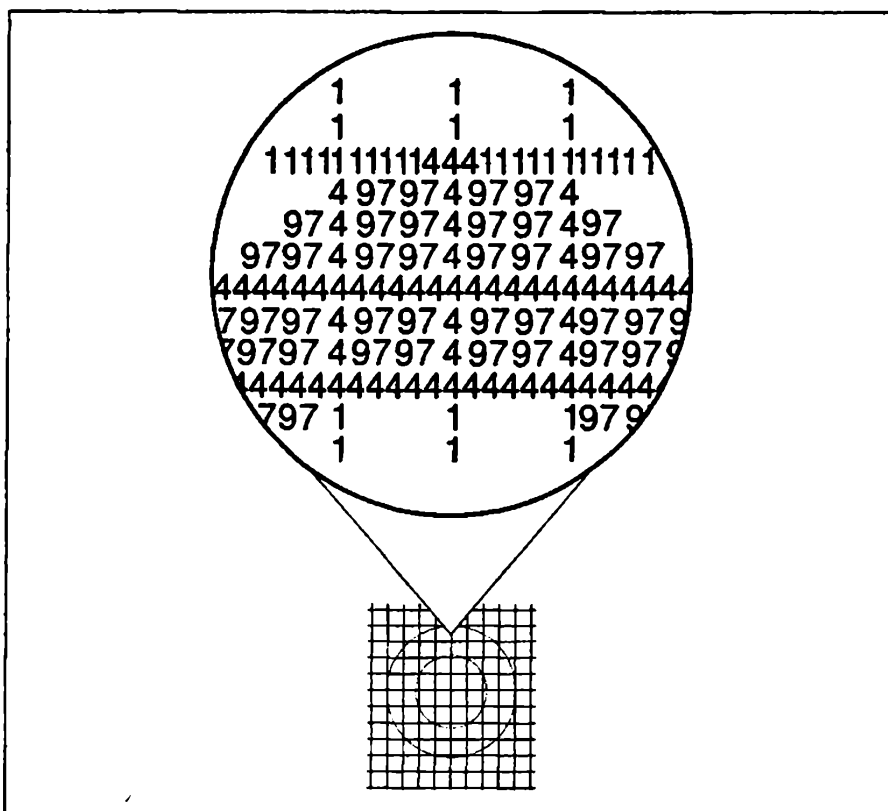
Un altro circuito cerebrale, specializzato nei colori ma non molto abile nelle forme e nelle collocazioni, arriva ad una discriminazione di colore (diciamo, rosa 97) con cui «etichetta» la zona circostante, e l'etichetta viene affibbiata all'intera regione.

Non è ancora ben chiaro perché queste particolari discriminazioni debbano verificarsi in queste condizioni, ma i dubbi riguardano i meccanismi causali che conducono a questa etichettazione errata della regione e non i successivi «prodotti» (se così si può dire) del sistema visivo. Ma non manca qualcosa? La mia spiegazione di come ad una particolare regione possa essere attribuita un'etichetta numerica che individua un colore non si è fermata troppo presto? Questa ricetta per ottenere un'immagine colorata non deve essere eseguita da qualche parte? Non si deve «riempire» l'anello con il rosa 97? Dopo tutto, potresti essere tentato di dire che vedi il rosa! Non vedi sicuramente una

11.10

regione delimitata con un numero scritto al centro. Il rosa che vedi non sta nel mondo esterno (non è un pigmento o una tinta o una «luce colorata»), quindi deve stare «qui dentro» – deve essere in altre parole un figmento rosa.

Dobbiamo fare attenzione a distinguere l'ipotesi del «figmento rosa» da altre che costituirebbero alternative legittime ad una spiegazione che si è fermata alla rappresentazione dei colori tramite numeri. Per esempio, potrebbe essere che in qualche parte del cervello ci sia una rappresentazione approssimativamente continua delle regioni colorate – una mappa di bit – tale che «ogni pixel» della regione debba essere etichettato «colore 97», più o meno come nella figura 11.11.

Questa è una possibilità empirica. Potremmo escogitare degli esperimenti per confermarla o refutarla. Dovremmo rispondere a questo interrogativo: nel cervello c'è un mezzo rappresentazionale nel quale il valore di un certo parametro variabile (l'intensità o qualsiasi altra codifica del colore) deve essere diffuso o ripetuto su tutti i pixel rilevanti di una matrice? Oppure c'è soltanto «un'unica etichetta» della regione senza che sia richiesto nessun ulteriore «riempimento» o «diffusione»? Quali tipi di esperimenti incoraggerebbero un modello del genere per l'effetto della diffusione al neon del colore? Sarebbe impressionante, per esempio, se si potesse mostrare che il colore in determinate circostanze si diffonde a poco a poco – colando

11.11

come sangue dalle linee centrali rosse e raggiungendo gradualmente i contorni soggettivi.[13] Non voglio anticipare una risposta, del resto ho sollevato questo problema solo per illustrare la mia affermazione che sebbene ci siano ancora molte questioni empiriche non risolte nella spiegazione del fenomeno della diffusione al neon del colore, nessuna di queste implica differenze sulla generazione o non generazione del figmento in una «decodificazione» del sistema neurale di codificazione.

Gli interrogativi che riguardano il modo in cui il cervello «riempie» non possono essere risolti tramite l'introspezione, perché, come abbiamo visto nel capitolo 4, l'introspezione ci fornisce – al soggetto, come allo sperimentatore «esterno» – solo il contenuto della rappresentazione e non le caratteristiche del mezzo rappresentazionale stesso. Per saperne di più sul mezzo, dobbiamo condurre ulteriori esperimenti.[14] Ma per alcuni fenomeni, possiamo già essere *abbastanza* sicuri che il mezzo di rappresentazione sia una versione di qualcosa di efficiente, come il colore tramite numeri, e non approssimativamente continuo, come la mappa di bit.

Si consideri come il cervello debba trattare la carta da parati, per esempio. Supponiamo di entrare in una stanza e di notare che le pareti siano ricoperte da una carta da parati che

raffigura una successione regolare di centinaia di barche a vela identiche o – per rendere omaggio a Andy Warhol – di ritratti fotografici identici di Marilyn Monroe. Per identificare una fotografia come il ritratto di Marilyn Monroe, dobbiamo far in modo che l'immagine cada nelle aree ad alta risoluzione delle nostre fovee retiniche.

Come vedemmo nell'esperimento con la carta da gioco nel capitolo 3, la nostra visione *parafoveale* (fornita dal resto della retina) non ha una risoluzione molto buona; non possiamo nemmeno identificare un fante di picche tenuto a distanza di braccio. Eppure sappiamo che se dovessimo entrare in una stanza con i muri ricoperti di foto identiche di Marilyn Monroe, «vedremmo istantaneamente» che le cose stanno effettivamente così. Vedremmo in una frazione di secondo che ci sono «centinaia e centinaia di ritratti identici, dettagliati e accurati di Marilyn Monroe».

Poiché in un secondo i nostri occhi compiono quattro o cinque movimenti saccadici, non più di una o due Marilyne possono essere entrate nelle nostre fovee nel tempo che ci occorre per saltare alla conclusione e *quindi a vedere* centinaia di identiche Marilyne. Sappiamo che la visione parafoveale *non riuscirebbe* a distinguere una Marilyn da una macchia simile a Marilyn, ma tuttavia ciò che vediamo *non* è una carta da parati con una Marilyn al centro contornata da varie e indistinte macchie simili a Marilyn.

Ora, non è possibile ipotizzare che il cervello prenda una delle immagini foveali ad alta risoluzione di Marilyn che ha ricevuto e la riproduca, come una fotocopiatrice, su una mappatura interna della superficie del muro? Questa sarebbe l'unica maniera in cui i dettagli ad alta risoluzione utilizzati per identificare Marilyn potrebbero «essere introdotti nello sfondo», poiché la visione parafoveale non è sufficientemente acuta per fornirli direttamente. Suppongo che sarebbe possibile in linea di principio, ma il cervello quasi certamente non deve affatto preoccuparsi di *questo* lavorio di riempimento! Avendo identificato una singola Marilyn, e non avendo ricevuto alcuna informazione che smentisca che le altre macchie non siano Marilyn, salta alla conclusione che anche le altre siano Marilyn, e etichetta l'intera regione «altre Marilyne» senza dover rendere affatto le Marilyne.[15]

Naturalmente a noi non sembra così. Ci sembra di vedere realmente centinaia di Marilyne identiche. E in un certo senso è vero: ci sono, infatti, centinaia di Marilyne identiche lì sul

muro e noi le stiamo guardando. Quello che non è vero, comunque, è che ci siano centinaia di Marilyne identiche rappresentate nel nostro cervello. Il nostro cervello si limita a rappresentare in qualche modo *che* ci sono centinaia di Marilyne identiche, e per quanto vivida possa essere la nostra impressione di vedere tutti questi dettagli, essi sono nel mondo, non nella nostra testa. E nessun figmento viene usato per rendere questa apparenza, perché l'apparenza non viene resa affatto, neppure come una mappa di bit.

Ora possiamo quindi rispondere alla nostra domanda sul punto cieco. Il cervello non deve «riempire» il punto cieco, poiché la regione in cui esso cade è già etichettata (per es., «a scacchi» o «Marilyne» o semplicemente «altre cose simili»). Se dovesse ricevere segnali contraddittori da qualche regione, il cervello abbandonerebbe o riaggiusterebbe la sua generalizzazione, ma non ricevere nessun segnale dalla regione del punto cieco non è equivalente a ricevere segnali contraddittori.

L'assenza di segnali di conferma dalla regione del punto cieco non rappresenta un problema per il cervello; poiché non ha mai ricevuto informazioni da quella lacuna nella retina, il cervello non ha sviluppato nessuna agenzia epistemicamente affamata che chiede di essere nutrita da quella regione. Tra tutti gli homunculi della visione, nemmeno uno ha il compito di coordinare le informazioni provenienti da quella regione dell'occhio, così, quando da lì non arriva alcuna informazione, nessuno si lamenta. L'area è semplicemente negletta. In altre parole, tutte le persone normalmente vedenti «soffrono» di questa piccola «anosognosia». Non siamo consapevoli del nostro «difetto» – del fatto che non stiamo ricevendo alcuna informazione visiva dai nostri punti ciechi. (Una buona rassegna sull'anosognosia è McGlynn e Schacter, 1989.)

Il punto cieco è un buco spaziale, ma ci possono essere anche buchi temporali. I più piccoli sono gli intervalli che si verificano quando i nostri occhi lanciano frecciate qui e là durante i movimenti saccadici. Noi non notiamo questi intervalli, ma non devono essere riempiti *perché* non siamo progettati per notarli. Gli analoghi temporali degli scotomi potrebbero essere le «assenze» che si verificano durante gli attacchi epilettici del piccolo male. Queste possono essere notate dagli ammalati, ma solo per inferenza: essi non possono «vederne i bordi» come noi non possiamo vedere i bordi dei nostri punti ciechi,

ma possono essere colpiti, retrospettivamente, dalla discontinuità negli eventi di cui hanno avuto esperienza.

La pecca fondamentale dell'idea del «riempimento» è che suggerisce che il cervello stia offrendo qualcosa quando invece sta ignorando qualcosa. E questo conduce anche i pensatori più sofisticati a commettere degli errori madornali, perfettamente compendiati da Edelman: «Uno fra i caratteri più sorprendenti della coscienza è la sua continuità» (1989, p. 149). Questo è completamente sbagliato. Uno fra i caratteri più sorprendenti della coscienza è la sua *dis*continuità – come è rivelato dal punto cieco e dagli intervalli saccadici, per prendere gli esempi più semplici. La discontinuità della coscienza è sorprendente proprio per l'*apparente* continuità della coscienza. Neumann (1990) sottolinea che la coscienza può essere in generale un fenomeno lacunoso, e, fintanto che i confini temporali delle lacune non siano sicuramente percepiti, non ci sarà alcun senso della lacunosità del «flusso» della coscienza. Come dice Minsky: «Nulla può *apparire* discontinuo se non ciò che è *rappresentato* come discontinuo. Paradossalmente, il nostro senso della continuità proviene dalla nostra meravigliosa *insensibilità* alla maggior parte dei cambiamenti, più che da un'autentica capacità di percezione» (1985, p. 502).

6. *La negligenza come perdita patologica dell'appetito epistemico*

Nel trattare con il punto cieco il motto del cervello potrebbe essere: non farmi domande e non ti dirò bugie. Come abbiamo visto nel capitolo 1, oltre a lenire la fame epistemica che incontra, il cervello non deve fare nient'altro. Ma cosa avviene quando c'è una fame epistemica molto minore di quella che dovrebbe esserci? Ci troviamo nelle situazioni descritte come patologie della negligenza (*neglect*).

Una delle forme più diffuse di negligenza è la *negligenza spaziale unilaterale*, in cui una parte del corpo, di solito la sinistra, è interamente negletta, a causa di un danno cerebrale nella parte opposta. Non solo la parte sinistra del corpo, ma anche tutto ciò che sta nell'immediata vicinanza del lato sinistro del corpo viene negletto. Se un gruppo di persone stanno in piedi attorno al letto di un paziente con negligenza spaziale unilaterale sinistra, egli si rivolgerà solo alle persone alla sua destra; se gli si chiede di contare le persone nella stanza, tenderà a trascurare le persone a sinistra e se qualcuno sulla sinistra cercherà

di attirare la sua attenzione, di solito non ci riuscirà. Eppure si può verificare che gli organi sensoriali del paziente continuano a recepire e ad analizzare e a reagire in vario modo agli stimoli che avvengono sulla sinistra. Cosa sta avvenendo nella testa del paziente? Il «lato sinistro dello spazio fenomenologico del soggetto è un vuoto»? Oppure «l'occhio della mente» del paziente non riesce a vedere il materiale che il cervello gli fornisce sulla parte sinistra del... palcoscenico del Teatro Cartesiano?

C'è una spiegazione più semplice, non in termini di rappresentazioni interne con curiose proprietà, ma in termini di *negligenza* in senso politico! È noto che Daniel Patrick Moynihan pensava che alcuni problemi impliciti nelle relazioni razziali in America si sarebbero risolti da soli se fossero stati trattati con una «negligenza benevola»! – se solo Washington e il resto del paese li avessero semplicemente ignorati per un po'. Non credo che il suo consiglio fosse buono, ma Moynihan aveva ragione su qualcosa: ci sono sicuramente delle circostanze in cui è necessaria una negligenza benevola – come nel modo in cui affrontiamo il problema del punto cieco.

Non ci sono homunculi – nell'immagine da me usata – che dovrebbero «occuparsi» dell'informazione proveniente dalla parte del campo visivo occupato dal punto cieco, così quando non arriva nulla, non c'è nessuno che si lamenta. Forse la differenza tra noi e coloro che soffrono di patologie come la negligenza, o altre forme di anosognosia, consiste nel fatto che sono stati uccisi quelli che si dovrebbero lamentare. Questa teoria è stata presentata, in termini meno coloriti, dal neuropsicologo Marcel Kinsbourne (1980), che chiama gli addetti interni alle lamentele «analizzatori corticali». Nei termini del modello che abbiamo sviluppato, la negligenza potrebbe essere descritta come una perdita d'influenza politica di alcuni partiti di folletti nel cervello, dovuta, in molti casi ma non tutti, alla morte o alla soppressione dei loro Deputati. Questi folletti sono ancora attivi, cercano di portare avanti le loro varie faccende e riescono talvolta ad avere perfino un qualche successo, ma non possono farcela più in certe competizioni contro coalizioni meglio organizzate.

In questo modello, la negligenza benevola dei nostri punti ciechi scivola quasi impercettibilmente nelle varie forme di negligenza moderatamente gravi di cui tutti soffriamo e anche nelle negligenze più stravaganti studiate dai neurologi. Per esempio, io stesso soffro di una forma piuttosto comune di negligenza. Non è molto seria, ma è tavolta imbarazzante, si

tratta della mia negligenza tipografica: sono patologicamente incapace di rilevare gli errori tipografici nelle bozze dei miei scritti, e riesco a rimediare a tale difficoltà solo a prezzo di uno strenuo sforzo di concentrazione. Non è, come suggerisce Baars, che il mio cervello «riempie» con l'ortografia corretta; non deve «riempire» poiché normalmente non dedica abbastanza attenzione a questi problemi da notare gli errori; la sua attenzione viene catturata da altri aspetti delle parole sulla pagina. Un'altra delle mie lievi disabilità è *la negligenza nei confronti dei compiti degli studenti*. È incredibile quanto possa diventare attraente la prospettiva di lavare il pavimento della cucina o cambiare la carta degli scaffali o mettere ordine nel mio libretto degli assegni quando ho una pila di compiti sulla mia scrivania che aspettano di essere valutati. Questo aspetto, l'interesse aumentato nei confronti delle alternative, è particolarmente evidente nella negligenza spaziale unilaterale; ad una prima approssimazione, più una cosa è sulla destra, più è degna di nota per un paziente con negligenza spaziale unilaterale sinistra. Forse la mia forma di negligenza più seria è, comunque, il mio caso disperato di *negligenza finanziaria*. Tenere la contabilità mi piace così poco, in realtà, che solo delle alternative davvero tremende, come la correzione dei compiti degli studenti, può forzare la mia attenzione sull'argomento. Questa negligenza ha serie conseguenze sul mio benessere, conseguenze che non è difficile farmi riconoscere, ma, a dispetto di questo allarmante e inutile richiamo alla mia sottostante razionalità, io continuo ad insistere nella mia negligenza, a meno che non vengano messe in atto drastiche misure di auto-manipolazione.

Non è che *non posso vedere* il mio libretto d'assegni, io *non lo voglio vedere*. E sebbene in momenti come questo, in cui sono calmo e riflessivo, possa riferire tutto ciò (dimostrando di non avere una *profonda* anosognosia sulle mie proprie disabilità), nel corso normale degli eventi parte di quello che io non noto è la mia negligenza nei confronti delle mie finanze. Una lieve anosognosia, in breve. Da questa prospettiva, l'unica cosa sorprendente sulle forme stravaganti di negligenza studiate dai neuropsicologi riguarda ciò che si neglige. Immagina qualcuno che neglige tutto ciò che sta *alla sua sinistra* (Bisiach e Luzzatti, 1978; Bisiach, 1988; Bisiach e Vallar, 1988; Calvanio, Petrone e Levine, 1987). O immagina qualcuno che ha perso la visione cromatica *ma non se ne lamenta* (Geschwind e Fusillo, 1966). O, anche, immagina qualcuno che è diventato cieco, ma non si è ancora accorto di questa grave perdita – sindrome di Anton o

rifiuto della cecità (Anton, 1899; McGlynn e Schacter, 1989, pp. 154-158).

Queste situazioni sono facilmente spiegabili in base alla teoria delle Molteplici Versioni della coscienza, giacché il Testimone centrale è stato sostituito da coalizioni di specialisti le cui specifiche fami epistemiche non possono essere immediatamente adottate da altri agenti se essi sono licenziati o in vacanza.[16] Quando queste fami epistemiche svaniscono, svaniscono senza tracce, lasciando il campo ad altre coalizioni, altri agenti con altre agende.

Ma lo stesso principio che spiega la negligenza fornisce uno scenario alternativo per la «mancanza di qualia visivi» nel nostro immaginario virtuoso della visione cieca. Avevo suggerito che una sua eventuale lamentela sull'assenza dei qualia *potrebbe* semplicemente essere interpretata come una sua constatazione della scarsità relativa dell'informazione ottenuta ora dalla sua visione e che tale constatazione venga però descritta erroneamente. Seguitavo facendo l'ipotesi che, se in un qualche modo si potesse accrescere la velocità in baud della raccolta dell'informazione, si potrebbe colmare almeno parte del divario tra questo tipo di visione e quella normale. Ora possiamo capire che un altro modo poco dispendioso per colmare lo stesso divario sarebbe quello di indebolire la sua fame epistemica o di ottundere in qualche modo la sua curiosità visiva. Dopo tutto, se nella sindrome di Anton una persona può essere completamente cieca e non accorgersene, un po' di negligenza strategicamente dislocata potrebbe trasformare il nostro soggetto che si lamenta sulla perdita dei qualia visivi in un soggetto che senza lamentele dichiara che la sua vista si è completamente ristabilita. Potrebbe sembrare che noi dovremmo saperne di più, ma è così? Mancherebbe qualcosa in tale persona? Non esiste alcun *figmento* nella visione normale, quindi non può essere il figmento ciò che manca. Cos'altro potrebbe essere?

7. *La presenza virtuale*

> Abbiamo la sensazione della realtà quando ogni domanda posta al nostro sistema visivo riceve una risposta così rapida da farci sembrare che quella risposta fosse già presente.
>
> Marvin Minsky (1985), p. 503

Ancora una volta, l'assenza di rappresentazione non è la stessa cosa che la rappresentazione dell'assenza. E la rappresentazione della presenza non è la stessa cosa che la presenza della rappresentazione. Ma questo è difficile da credere. Noi siamo intuitivamente convinti di avere in qualche modo una *conoscenza diretta* delle proprietà e delle caratteristiche speciali della nostra esperienza, e tale robusta convinzione deve essere affrontata da chiunque voglia sviluppare una buona teoria della coscienza. Io ho cercato di scalfire, di minare la sua autorevolezza, ma c'è ancora molto da fare. Otto può ancora tentare un'altra mossa:

La tua argomentazione sulle Marilyne nella carta da parati è in realtà una difesa indiretta del dualismo. Tu sostieni, in modo molto persuasivo, che nel cervello non ci sono centinaia di Marilyne ad alta risoluzione e concludi quindi che non stanno da nessuna parte! Ma poiché *ciò che noi vediamo* sono centinaia di Marilyne ad alta risoluzione, allora, *poiché*, come hai sostenuto, non sono in nessun luogo dei nostri cervelli, devono essere da qualche altra parte – nelle nostre menti non fisiche!

Le centinaia di Marilyne sulla carta da parati sembrano essere presenti nella nostra esperienza, sembrano essere nella nostra mente, non solo sul muro. Ma poiché, come sappiamo, il nostro sguardo si può spostare in una frazione di secondo per ottenere informazioni da qualsiasi punto del nostro ambiente visivo, perché mai il cervello si dovrebbe preoccupare di portare dentro di sé fin dall'inizio tutte quelle Marilyne? Perché non lascia che sia il mondo ad immagazzinarle, gratis, per lui, fino a che non gli servono?

Paragoniamo il cervello a una biblioteca. Alcune biblioteche di istituti di ricerca sono dei giganteschi magazzini, che contengono milioni di libri, tutti negli appositi scaffali e accessibili in modo abbastanza facile. Altre biblioteche tengono meno libri a portata di mano, ma dispongono di un sistema di accessione generoso ed efficiente: comprano tutti i libri che vengono richiesti dagli utenti o se li fanno prestare da altre biblioteche, usando un rapido sistema di prestito interbibliotecario. Se non tieni i libri accatastati nell'edificio, i ritardi nelle consegne sono maggiori, ma non molto maggiori. Possiamo immaginare un sistema elettronico di prestito interbibliotecario (che utilizza fax o *file* di calcolatori) che possa consegnare un libro ottenuto dall'esterno più velocemente di uno prelevato dai propri scaffali da un rapidissimo bibliotecario. Un informatico potrebbe dire che i

libri di questo sistema sono sempre «virtualmente presenti» nella biblioteca o che la biblioteca possiede una «collezione virtuale» cento o mille volte più grande della sua reale collezione cartacea.

Ora, come potremmo sapere noi, come Utenti delle nostre biblioteche cerebrali, quali delle cose di cui ci serviamo sono sempre state *lì* e quali sono state ottenute dal cervello tramite rapide sortite nel mondo esterno alla caccia d'informazioni? Accurati esperimenti, condotti secondo il metodo dell'eterofenomenologia, possono rispondere a questo interrogativo, ma l'introspezione, da sola, non può riuscirci. Il che non ci impedisce, comunque, di pensare di poterci riuscire. In assenza di qualunque prova in un senso o nell'altro, la nostra tendenza naturale è di saltare alla conclusione che molto di più è presente. Io ho chiamato questa tendenza la Trappola Introspettiva (Dennett, 1969, p. 179) e Minsky la chiama l'Illusione dell'Immanenza: «Ogni volta che sappiamo rispondere a una domanda senza ritardo apprezzabile, si ha l'impressione che questa risposta fosse già attiva nella nostra mente» (Minsky, 1985, p. 298).

Quella del sistema di prestito interbibliotecario è un'analogia utile ma incompleta, perché il cervello non possiede semplicemente una strumentazione per acquisire informazioni su qualunque argomento esterno ci possa interessare; ha anche milioni di sentinelle che scrutano quasi in continuazione una porzione del mondo esterno, pronte a suonare l'allarme e ad *attirare* la nostra attenzione su qualunque avvenimento nuovo e importante nel mondo. Nella visione, ciò viene realizzato dai bastoncelli e dai coni parafoveali delle nostre retine e dagli agenti neurali collegati a queste sentinelle che sono specializzati nel rilevare il cambiamento e il movimento. Se uno di questi agenti suona l'allarme – «Un cambiamento nel mio settore!» – ciò scatena quasi istantaneamente un movimento saccadico, che porta la fovea a centrare la regione interessante, così le novità possono essere localizzate, identificate e affrontate. Il sistema di allarme è così affidabile che è difficile introdurre furtivamente qualche cambiamento nel mondo visibile senza che l'intero sistema visivo ne sia informato, ma, con l'aiuto di trucchi ad alta tecnologia, le sentinelle possono essere aggirate, con sorprendenti risultati.

Quando i nostri occhi saltellano qua e là nei movimenti saccadici, le contrazioni muscolari che fanno roteare i bulbi oculari sono azioni *balistiche*: i punti di fissazione sono missili *non*

guidati le cui traiettorie al decollo determinano il punto e il momento in cui colpiranno il nuovo bersaglio. Per esempio, se stai leggendo un testo sullo schermo di un calcolatore, i tuoi occhi faranno un salto lungo poche parole a ogni movimento saccadico – quanto più si è buoni lettori, tanto più lungo e più rapido sarà il salto. Che cosa si proverebbe se un mago, una sorta di demone maligno cartesiano su scala ridotta, potesse cambiare il mondo durante quei pochi millisecondi che i nostri occhi impiegano a saltare alla loro prossima destinazione? È incredibile, ma un calcolatore dotato di un tracciatore automatico dell'occhio può rilevare e analizzare il decollo nei primi pochi millisecondi di un movimento saccadico, calcolare quale sarà il nuovo punto d'impatto e, *prima che il movimento saccadico sia finito*, cancellare la parola che occupa il punto d'impatto e sostituirla con una parola differente della stessa lunghezza. Cosa vedrai? Solo la nuova parola, e senza alcun senso di cambiamento. Mentre osservi con attenzione il testo sullo schermo, ti sembrerà che tutto il mondo sia stabile come se le parole fossero scolpite nel marmo, ma a qualcun altro che leggesse lo stesso testo sopra le tue spalle (e avesse dei movimenti saccadici a un ritmo differente) lo schermo sembrerebbe un guazzabuglio in movimento.

L'effetto è irresistibile. Quando mi imbattei per la prima volta nell'esperimento del tracciatore oculare, e fui stupito dell'incapacità dei soggetti a notare i cambiamenti su quello schermo svolazzante, chiesi agli sperimentatori se mi potevano far provare. Volevo rendermi conto di persona. Fui fatto accomodare di fronte all'apparato, la mia testa era immobilizzata perché dovevo mordere una barra; questo rendeva le cose più facili al tracciatore, che doveva far rimbalzare un fascio di luce impercettibile sul cristallino dell'occhio del soggetto e poi analizzare il fascio riflesso per rilevare i movimenti dell'occhio. Mentre attendevo che gli sperimentatori mettessero in funzione l'apparato, lessi il testo sullo schermo. Aspettai, e aspettai, ansioso che la prova cominciasse. Divenni impaziente. «Perché non l'accendete?», chiesi. «È acceso», mi risposero.

Poiché tutti i cambiamenti sullo schermo avvengono durante i movimenti saccadici, le sentinelle non riescono a dare nessun allarme effettivo. Fino a poco tempo fa questo esperimento era chiamato la «soppressione dei movimenti saccadici». L'idea era che il cervello deve in qualche modo bloccare i segnali in ingresso dagli occhi durante i movimenti saccadici, poiché nessuno può notare i cambiamenti che avvengono nel

campo visivo durante i movimenti saccadici e, naturalmente, nessuno si lamenta di cambiamenti allarmanti e turbinanti. Ma un astuto esperimento con un tracciatore oculare (Brooks e colleghi, 1980) ha mostrato che se uno stimolo – come una parola o una lettera dell'alfabeto – viene mosso in sincronia con il movimento saccadico, tenendo il passo dell'«ombra» della fovea mentre corre verso il suo nuovo punto di atterraggio, viene visto e identificato facilmente dal soggetto. I segnali in ingresso dall'occhio non vengono bloccati lungo la strada che li porta al cervello durante i movimenti saccadici, ma in condizioni normali non sono utilizzabili – tutto scorre troppo velocemente per avere un senso – quindi il cervello tratta tutto ciò con benevola negligenza. Se tutte le sentinelle inviano i loro allarmi contemporaneamente, la cosa migliore da fare è semplicemente ignorarle.

Nella situazione sperimentale in cui mi sono trovato, le parole sullo schermo venivano cancellate e sostituite durante i miei movimenti saccadici. Se la tua visione parafoveale non può discriminare la parola nel punto d'impatto prima che il movimento saccadico vi arrivi, una volta che sei lì e la identifichi non ci può essere nessuna registrazione o memoria precedente nel cervello con cui confrontarla. Lo scambio non può essere notato perché l'informazione logicamente richiesta per poterlo notare semplicemente *non è lì*. Naturalmente *ti sembra* mentre leggi questa pagina che tutte le parole sulla riga sono in un certo senso presenti nella tua coscienza (sullo sfondo) anche prima che tu ti rivolga specificamente ad esse, ma questa è un'illusione. Esse sono presenti solo virtualmente.

Nel cervello ci sono naturalmente *alcune* informazioni sulle parole circostanti – sufficienti ad aver guidato e istigato il più recente movimento saccadico, per esempio. Esattamente quale informazione è già *lì*? Gli esperimenti con i tracciatori e simili apparati possono determinare i limiti di ciò che si può notare, e quindi determinare i limiti di ciò che è *presente* nella nostra mente. (Vedi, per es., Pollatsek, Rayner e Collins, 1984; Morris, Rayner e Pollatsek, 1990.) Insistere, come era tentato di fare Otto, che ciò che non è *lì* nel cervello deve tuttavia essere *lì* nella mente perché *sembra* senz'altro essere lì, è insensato. Perché, come abbiamo appena visto, non starebbe «lì» in alcun senso che possa fare una differenza *per l'esperienza personale di Otto*, per non parlare della sua capacità di superare dei test, premere dei pulsanti e così via.

A questo punto Otto, il nostro critico, pretende una ricapitolazione, giacché si sente ingannato in qualche punto dell'argomentazione. In un dialogo fra lui e me affronteremo i punti che gli stanno a cuore, sperando di risolvere nello stesso tempo anche una buona parte dei tuoi dubbi. Otto comincia:

Mi sembra che hai negato l'esistenza dei fenomeni più reali, più indubitabili: quel reale sembrare di cui nemmeno Cartesio nelle sue Meditazioni *ha potuto dubitare.*

In un certo senso, hai ragione; questo è proprio quello di cui sto negando l'esistenza. Ma ritorniamo al fenomeno della diffusione al neon del colore. Sembra che sulla copertina ci sia un lucente anello rosa.

Sembra proprio così.

Ma non c'è nessun lucente anello rosa. Non in realtà.

Giusto. Ma sembra proprio che ci sia!

Giusto.

Quindi dove sta?

Dove sta, che cosa?

Il lucente anello rosa.

Non c'è nessun lucente anello rosa; credevo che l'avessi appena ammesso.

Be', sì, non c'è nessun lucente anello rosa lì sulla pagina, ma sembra proprio che ci sia.

Giusto. Sembra che ci sia un lucente anello rosa.

Quindi, parliamo di quell'*anello.*

Quale?

Quello che sembra esserci.

Non esiste una cosa come un anello rosa che sembra solamente esserci.

Guarda che io non dico *soltanto che sembra esserci un lucente anello rosa;* sembra realmente *esserci un lucente anello rosa!*

Sono perfettamente d'accordo. Non ho mai voluto accusarti di parlare a vanvera! Tu intendi realmente ciò che dici quando dici che sembra esserci un lucente anello rosa.

Guarda che io non lo intendo soltanto. Io non penso *soltanto che sembra esserci un lucente anello rosa;* realmente sembra esserci *un lucente anello rosa!*

Ora l'hai detta grossa. Sei caduto in una trappola, assieme a tanti altri. Non penserai davvero che ci sia una differenza tra il credere (pensare, giudicare, decidere, essere della più sincera opinione) che qualcosa ti sembra rosa e il credere che qualcosa *realmente ti sembra* rosa. Una tale differenza non esiste. Non esiste il fenomeno del *sembrare realmente* – oltre il fenomeno di giudicare in un senso o nell'altro che qualcosa è in un certo modo.

Ricordati della carta da parati alle Marilyne. Il muro è, in effetti, ricoperto di Marilyne ad alta risoluzione. Inoltre, a te sembra proprio così! Ti sembra che il muro sia ricoperto di Marilyne ad alta risoluzione. Beato te, il tuo apparato visivo ti ha condotto a una credenza veritiera su un aspetto del tuo ambiente circostante. Ma nel tuo cervello (o nella tua mente) non sono rappresentate innumerevoli Marilyne che *realmente sembrano*. Non esiste un mezzo che *riproduce* i dettagli della carta da parati, che li *presenta* al tuo Testimone interiore. Tutto quello che accade in questo caso è che ti sembra che ci siano innumerevoli Marilyne ad alta risoluzione lì sul muro (e questa volta hai ragione, ci sono davvero). Altre volte ti puoi sbagliare; per esempio nel fenomeno phi colorato, ti può sembrare che un unico punto luminoso si è mosso e ha cambiato colore a metà strada, mentre in realtà c'erano soltanto due punti luminosi differentemente colorati che lampeggiavano. Questo sembrarti così non richiede una *presentazione* nel cervello, non più di quanto i giudizi cromatici del cervello, una volta formulati, abbiano bisogno di essere successivamente *decodificati* in qualche modo.

Ma allora che cosa si verifica quando mi sembra che ci sia un lucente anello rosa? Qual è la spiegazione precisa offerta dalla tua teoria? Mi sembri terribilmente evasivo a questo riguardo.

Sospetto che tu abbia ragione. È tempo di fare chiarezza e di presentare una spiegazione precisa, ma confesso che mi sento forzato a cominciare con una caricatura, per poi rivederla. Non riesco a scoprire un modo più diretto per spiegarmi.

L'ho notato. Vai avanti.

Supponi che ci *sia* un Autore Centrale. Ma supponi che invece di sedere nel Teatro Cartesiano a vedere la Presentazione, l'Autore Centrale stia seduto nel buio e abbia presentimenti – all'improvviso gli capita di pensare che ci sia qualcosa di rosa lì fuori, così come ti può capitare di pensare all'improvviso che qualcuno ti segue.

Che cosa sono i presentimenti, esattamente? Di che cosa sono fatti?

Questa è un'ottima domanda, alla quale però posso rispondere per il momento solo in modo evasivo, nella caricatura. Questi presentimenti sono proposizioni che l'Autore Centrale rivolge a se stesso nel suo linguaggio speciale, il mentalese. Così la sua vita consiste di una sequenza di *giudizi*, che sono frasi in mentalese, che esprimono una proposizione dopo un'altra, a velocità enorme. Alcune di queste egli decide di pubblicarle, nella loro traduzione in linguaggio naturale.

Questa teoria ha la virtù di fare a meno del figmento, della proiezione nello spazio fenomenico, del riempimento di tutti gli spazi vuoti dello Schermo Teatrale, ma ha ancora un Autore Centrale e un Linguaggio del Pensiero. Quindi rivediamo la teoria. Come prima cosa, sbarazziamoci dell'Autore Centrale distribuendo su tutto il cervello, nello spazio e nel tempo, i suoi giudizi – ogni atto che serve a fare una discriminazione, un discernimento, a fissare un contenuto, avviene in qualche luogo, ma non c'è un Discriminatore che svolge tutto il lavoro. Poi, sbarazziamoci del Linguaggio del Pensiero; il contenuto dei giudizi non deve essere esprimibile in una forma «proposizionale» – questo è un errore, un caso di proiezione errata, perché eccessiva, delle categorie del linguaggio sulle attività del cervello.

Così i presentimenti sono come atti linguistici tranne che non c'è né Attore né Discorso!

Be', sì. Ciò che c'è, in realtà, sono solo vari eventi di fissazione del contenuto che si verificano in vari punti e vari tempi nel cervello. Questi non sono gli atti linguistici di qualcuno, e quindi non devono essere formulati in un linguaggio, ma sono abbastanza *simili* agli atti linguistici; hanno un contenuto e hanno l'effetto di informare vari processi di questo contenuto. Abbiamo già considerato in modo più dettagliato questo aspetto nei capitoli 5-10. Alcuni di questi atti in cui un contenuto viene fissato hanno effetti ulteriori, che alla fine conducono alla pronuncia di frasi – nel linguaggio naturale – o pubblicamente o solamente in privato. E così nasce un testo eterofenomenologico. Quando viene interpretato, si crea l'illusione benevola che ci sia un Autore. Questo è sufficiente a produrre l'*etero*fenomenologia.

*E l'*effettiva *fenomenologia?*

Non esiste una cosa del genere. Ricordati della discussione sull'interpretazione della narrativa. Quando ci imbattiamo in un romanzo che è vagamente autobiografico, scopriamo di poter associare gli eventi fittizi a molti eventi reali della vita dell'autore, cosicché in un certo senso possiamo dire che il romanzo *parla di* quegli eventi reali. L'autore può anche non accorgersene affatto, ma tuttavia, in questo senso un po' forzato, è vero; il testo parla di quegli eventi, perché quelli sono gli eventi reali che spiegano perché *questo* testo è stato creato.

Ma di che parla il testo nel senso non *forzato?*

Di nulla. È letteratura. *Sembra che si riferisca* a vari personaggi, luoghi ed eventi, ma questi eventi non sono mai accaduti; non si riferisce *realmente* a nulla.

Ma quando leggo un romanzo, questi fatti inventati diventano vivi! Qualcosa si verifica in me; io visualizzo *gli eventi. L'atto di leggere e interpretare un testo come un romanzo* crea *delle cose nuove nella mia immaginazione: immagini di personaggi e imprese. Dopo tutto, quando andiamo a vedere un film tratto da un romanzo che abbiamo letto, spesso pensiamo: «La protagonista non è affatto come l'avevo immaginata!».*

D'accordo. In *Fearing Fictions*, il filòsofo Kendall Walton (1978) sostiene che questi atti di immaginazione da parte di un interprete integrano il testo nella stessa maniera delle immagini di un'edizione illustrata di un romanzo, «si combinano con il romanzo per formare un mondo [fittizio, eterofenomenologico] più "ampio"» (p. 17). Queste aggiunte sono perfettamente reali, ma sono soltanto altro «testo» – fatto non di figmento, ma di giudizi. Nella fenomenologia non c'è altro che questo.

Ma sembra esserci altro!

Esattamente! *Sembra esserci la fenomenologia.* Questo è un fatto che un eterofenomenologo concede volentieri. Ma da questo fatto innegabile e universalmente attestato *non* segue che *esista realmente la fenomenologia.* Questo è il punto.

Stai negando allora che la coscienza sia un qualcosa di pieno?

Sì, infatti. Questo fa parte di ciò che sto negando. La coscienza è lacunosa e sparsa, e non contiene neanche la metà delle cose che la gente pensa!

Ma, ma...

Ma la coscienza sembra senz'altro qualcosa di pieno?

Sì!

Sono d'accordo; sembra essere qualcosa di pieno; e sembra perfino un «fatto sorprendente» sulla coscienza che sia continua, come dice Edelman, ma...

Lo so, lo so: dal fatto che sembra *che sia un pieno non segue che* sia *un pieno.*

Ora ci sei.

Ma io ho un altro problema con questa stanza degli specchi che chiami una teoria. Tu dici che è solo come se *ci fosse un Autore Centrale,* come se *ci fosse un unico Comandante,* come se *ci fosse un luogo dove tutto converge! Non capisco questa politica del* come se!

Forse un altro esperimento di pensiero te la renderà più digeribile. Immagina di visitare un altro pianeta e di scoprire che gli scienziati del luogo abbiano formulato una teoria piuttosto affascinante: ogni entità fisica ha dentro di sé un'anima e ogni anima ama un'altra anima; quindi, le entità tendono a muoversi una verso l'altra, spinte dall'amore che la loro anima interiore sente per le altre. Possiamo supporre, inoltre, che questi scienziati abbiano sviluppato dei sistemi molto accurati di localizzazione delle anime, così che, avendo determinato la localizzazione precisa nello spazio fisico dell'anima di un'entità, possano rispondere a domande sulla sua stabilità («Cadrà perché la sua anima è troppo in alto»), sulla vibrazione («Se metti un oggetto controbilanciante, con un'anima abbastanza grande, sul lato di quella ruota motrice, le oscillazioni si regolarizzeranno»), e su molte altre questioni tecniche.

Quello che potremmo dir loro, naturalmente, è che si sono imbattuti nel concetto di centro di gravità (o, per essere più accurati, centro di massa), e stanno solo trattandolo un po' troppo cerimoniosamente. Potremmo dir loro che possono benissimo continuare a parlare e pensare nel modo in cui facevano – tutto quello che dovrebbero fare è di abbandonare un bagaglio metafisico non necessario. C'è un'interpretazione più semplice, più austera (e molto più soddisfacente) degli stessi fatti che vengono compresi con la loro fisica animistica. Ci chiedono: esistono le anime? Be', certo, rispondiamo – è solo che sono degli *abstracta*, astrazioni matematiche invece di pepite di sostanze misteriose. Sono finzioni estremamente utili. È *come se* ogni oggetto attraesse ogni altro oggetto concentrando tutte le sue attrattive gravitazionali in un singolo punto – ed è molto più facile calcolare il comportamento di un sistema usando questa finzione ben strutturata che scendere in una marea di dettagli confusi – ogni punto che attrae ogni altro punto.

Mi sento come se *mi avessero appena derubato.*

Non dire che non ti avevo avvertito. Non puoi aspettarti che la coscienza si riveli essere *proprio come* la volevi. A parte questo, a cosa stai davvero rinunciando?

Solo alla mia anima.

Non in un senso coerente e difendibile. Tutto quello a cui stai rinunciando è una pepita speciale che però non potrebbe dav-

vero essere speciale. Perché avresti un concetto migliore di te stesso se risultassi essere una specie di mente-perla in un cervello-ostrica? Cosa ci sarebbe di così speciale nell'essere una mente-perla?

Una mente-perla potrebbe essere immortale, a differenza di un cervello.

L'idea che il Sé – o l'Anima – sia in realtà solo un'astrazione sembra a molte persone solo un'idea negativa, una negazione piuttosto che un qualcosa di positivo. Ma in realtà ha molto a suo favore, inclusa – se ti può interessare – una versione teoreticamente più solida dell'immortalità potenziale rispetto a qualunque altra che possa essere trovata nelle idee tradizionali di anima, ma per questo dobbiamo aspettare fino al capitolo 13. Prima dobbiamo affrontare definitivamente i qualia, che *ancora* hanno una presa sulla nostra immaginazione.

12

QUALI QUALIA?

1. *Un filo nuovo per il nostro aquilone*

> Gettato in un intervallo causale, un quale non potrà fare altro che passarci attraverso.
>
> IVAN FOX (1989), p. 82

Quando il filo dell'aquilone si aggroviglia, è sempre possibile, in linea di principio, riuscire a sbrogliarlo, soprattutto se siamo pazienti e analitici. Ma c'è un punto oltre il quale ciò che è valido in principio cede il passo, e la praticità trionfa: è meglio andare a comprare un filo nuovo. Ci costa davvero di meno alla fine di tutto il lavorio necessario per salvare quello vecchio, e riusciremo a far volare il nostro aquilone molto prima. La stessa cosa è valida, secondo me, per la discussione filosofica sui qualia, un tormentato groviglio, sempre più involuto e bizzarro, di esperimenti ideali, espressioni gergali, facezie, allusioni a presunte refutazioni, risultati «assodati» che dovrebbero essere rinviati al mittente e una miriade di scansafatiche e perdigiorno. Da certi pasticci è meglio stare alla larga, e quindi non mi appresterò ad un esame analitico di questa letteratura, anche se contiene momenti di profondità e ingegno di cui ho beneficiato (Shoemaker, 1975, 1981, 1988; White, 1986; Kitcher, 1979; Harman, 1990; Fox, 1989). In passato ho tentato di sbrogliare la matassa (Dennett, 1988a), ma ora penso che sia meglio provare a ripartire più o meno da capo.

Non è difficile capire perché i filosofi hanno finito per ingarbugliarsi in questi nodi sui qualia. Sono partiti da dove sarebbe partita qualunque persona di giudizio: dalle loro intuizioni più forti e più chiare sulle loro menti. Quelle intuizioni, purtroppo, formano un cerchio chiuso di dottrine che si sorreggono mutuamente e imprigionano la loro immaginazione nel Teatro Cartesiano. Sebbene i filosofi abbiano scoperto i para-

dossi inerenti a questo cerchio chiuso di idee – ecco perché la letteratura sui qualia esiste – essi non avevano una *prospettiva completa e alternativa* su cui saltare, e quindi, fidandosi delle loro intuizioni ancora molto forti, sono stati ritrascinati nella loro prigione paradossale. Ecco perché la letteratura sui qualia diventa sempre più involuta, invece di produrre un qualche accordo. Ma ora abbiamo messo in piedi proprio una prospettiva alternativa, il modello delle Molteplici Versioni. Usandolo possiamo offrire una spiegazione chiara e piuttosto differente dei problemi. Poi possiamo fare una pausa nei paragrafi 4 e 5 e paragonarlo alla posizione che spero rimpiazzerà.

Un eccellente libro introduttivo sul cervello contiene il seguente brano:

> Il «colore» come tale non esiste nel mondo; esiste solo nell'occhio e nel cervello dell'osservatore. Gli oggetti riflettono la luce di varie lunghezze d'onda, ma queste onde luminose in sé non hanno colore. (Ornstein e Thompson, 1984, p. 55)

Questo è un buon tentativo di esprimere il sapere comune, ma si deve notare che, preso in senso stretto e letterale, non può essere ciò che gli autori intendono, e non può essere vero. Il colore, essi dicono, non esiste «nel mondo», ma solo «nell'occhio e nel cervello» dell'osservatore. Ma l'occhio e il cervello dell'osservatore sono nel mondo, fanno parte del mondo fisico come gli oggetti visti dall'osservatore. E, come quegli oggetti, l'occhio e il cervello sono colorati. Gli occhi possono essere blu o marrone o verdi, e anche il cervello non è fatto *solo* di materia grigia (e bianca): oltre la *substantia nigra* c'è il *locus caeruleus*. Ma naturalmente i colori che stanno «nell'occhio e nel cervello dell'osservatore» in *questo* senso non sono quelli di cui parlano gli autori. Che cosa ci fa pensare che ci siano colori in qualche altro senso?

La scienza moderna – così si dice normalmente – ha rimosso il colore dal mondo fisico e lo ha sostituito con radiazioni elettromagnetiche incolori di varie lunghezze d'onda, che rimbalzano su superfici che riflettono e assorbono in varie maniere quella radiazione. Può sembrare che il colore sia *lì fuori*, ma non è così. È *qui dentro* – «nell'occhio e nel cervello dell'osservatore». (Se gli autori del brano non fossero dei materialisti rigorosi, avrebbero detto probabilmente che il colore sta nella *mente* dell'osservatore, salvandosi dalla sciocca interpretazione che abbiamo appena scartato, ma creandosi alcuni problemi anche

peggiori.) Ma, se non c'è un *figmento* interno che possa essere colorato in qualche speciale senso soggettivo, mentale, fenomenico, sembra che i colori siano destinati a sparire del tutto! *Qualcosa* devono pur essere i colori che conosciamo e amiamo, i colori che mescoliamo e accoppiamo. Dove, dove mai possono essere?

Questo che dobbiamo affrontare ora è un antico rompicapo filosofico. Nel diciassettesimo secolo, il filosofo John Locke (e, prima di lui, lo scienziato Robert Boyle) chiamò proprietà come i colori, gli aromi, i gusti e i suoni le *qualità secondarie*, per distinguerle dalle qualità primarie: dimensione, forma, movimento, numero e consistenza. Le qualità secondarie non erano di per sé «cose nella mente», ma piuttosto i *poteri delle cose* nel mondo (grazie alle loro particolari qualità primarie) di produrre o provocare certe cose nelle menti di osservatori normali. (E se non ci fossero osservatori nei paraggi? Questo è l'enigma eternamente popolare sull'albero che cade nella foresta. Fa rumore? La risposta viene lasciata come esercizio al lettore.) Il modo lockiano di definire le qualità secondarie è diventato parte della usuale interpretazione della scienza dell'uomo della strada; ha le sue virtù, ma è anche costretta a cedere degli ostaggi: le cose prodotte nella mente. La qualità secondaria *rosso*, per esempio, era per Locke il potere o la proprietà disposizionale di certe superfici degli oggetti fisici, grazie alle caratteristiche microscopiche della loro tessitura, di produrre in noi *l'idea del rosso* ogni qual volta la luce fosse riflessa da quelle superfici nei nostri occhi. Il potere degli oggetti esterni è abbastanza chiaro, almeno sembra, ma che razza di cosa è l'idea del rosso? È in qualche senso *colorata*, come qualcosa che indossa una bellissima toga blu? Oppure è, come una interessantissima discussione sul viola, solamente *su* un colore, senza essere essa stessa colorata? Ciò apre delle possibilità, ma come fa un'idea a essere solamente *su* un colore (per es., il colore rosso) se nulla da nessuna parte è rosso?

Che cosa è il rosso, in ogni modo? Che cosa sono i colori? I colori sono sempre stati l'esempio favorito dai filosofi e io seguirò la tradizione per il momento. Il problema principale nella tradizione emerge chiaramente dall'analisi filosofica di Wilfrid Sellars (1963, 1981b), che distingue le proprietà disposizionali degli oggetti (le qualità secondarie di Locke) da ciò che egli chiama le proprietà evenienti. Un cubetto di ghiaccio rosa nel frigorifero con la luce spenta ha la qualità secondaria rosa, ma non c'è nessuna presenza della proprietà *rosa eveniente* fino a

che un osservatore non apre lo sportello e guarda. Il rosa eveniente è una proprietà di qualcosa nel cervello o di qualcosa «nel mondo esterno»? In entrambi i casi, insiste Sellars, il rosa eveniente è una proprietà «omogenea» a qualcosa di reale. Parte di ciò che egli intende negare con questa insistenza sull'omogeneità sarebbe l'ipotesi che il rosa eveniente sia qualcosa come una *attività neurale di intensità 97 nella regione 75* del cervello. Egli intende negare anche che il mondo soggettivo della fenomenologia cromatica sia esaurito da qualcosa di così incolore come un *giudizio* che questa o quella cosa sia, o sembri essere, rosa. Per esempio, l'atto di richiamare con l'occhio della mente il colore di una banana matura e giudicare che il suo colore è giallo non porterebbe all'esistenza un caso di giallo eveniente (Sellars, 1981; Dennett, 1981b). Questo sarebbe meramente giudicare che qualcosa è giallo, un fenomeno che sarebbe di per sé privo di giallo eveniente come lo sarebbe una poesia sulle banane.

Sellars si è spinto fino ad affermare che tutte le scienze fisiche dovrebbero essere rivoluzionate per fare spazio al rosa eveniente e ai suoi simili. Pochi filosofi lo hanno seguito su questa strada così radicale, ma una versione di questo punto di vista è stata fatta risorgere recentemente da Michael Lockwood (1989). Altri filosofi, come Thomas Nagel, hanno supposto che perfino delle scienze rivoluzionate sarebbero incapaci di trattare con tali proprietà:

> I caratteri soggettivi dei processi mentali coscienti – in quanto opposti alle loro cause ed effetti fisici – non possono essere catturati dalla forma purificata di pensiero che è adatta a trattare con il mondo fisico che sottostà alle sue apparenze. (1986, p. 15)

I filosofi hanno adottato vari nomi per le cose nell'osservatore (o proprietà nell'osservatore) che avrebbero dovuto fornire una dimora sicura per i colori e per le altre proprietà che sono state bandite dal mondo «esterno» dai trionfi della fisica: «sensazioni grezze», «percetti», «qualità fenomeniche», «proprietà intrinseche dell'esperienza cosciente», «il contenuto qualitativo degli stati mentali» e, naturalmente, «qualia», il termine che userò. Ci sono differenze sottili nel modo in cui questi termini sono stati definiti, ma io non me ne curerò. Nel capitolo precedente sembrava che io volessi negare l'esistenza di *qualsiasi* proprietà del genere, e una volta tanto ciò che sembra così è

così. *Nego* che esistano tali proprietà. Ma (di nuovo torna questo tema) concedo di tutto cuore che sembra che i qualia esistano.

Sembra che ci siano i qualia, perché sembra davvero che la scienza ci abbia mostrato che i colori non possano essere lì fuori e quindi devono essere qui dentro. Inoltre, sembra che ciò che sta qui dentro non può essere *solo* il giudizio che formuliamo quando le cose ci sembrano colorate. Questo ragionamento è confuso, però. Quello che la scienza ci ha davvero mostrato è solo che le proprietà di riflessione della luce degli oggetti spingono le creature a entrare in vari stati discriminativi, distribuiti nei loro cervelli, che sottendono una schiera di disposizioni innate e abitudini apprese di varia complessità. E quali sono le *loro* proprietà? Qui possiamo giocare la carta di Locke per la seconda volta: questi stati discriminativi del cervello dell'osservatore hanno varie proprietà «primarie» (le loro proprietà meccaniche dovute alle loro connessioni, gli stati di eccitazione dei loro elementi, ecc.), e, in virtù di queste proprietà primarie, hanno varie proprietà secondarie, meramente disposizionali. Nelle creature umane con linguaggio, per esempio, questi stati discriminativi spesso finiscono per disporre le creature a esprimere giudizi verbali che alludono al «colore» di varie cose. Quando qualcuno dice «Lo so che l'anello non è realmente rosa, ma sembra proprio rosa», la prima parte esprime un giudizio su qualcosa nel mondo, la seconda esprime un giudizio del secondo ordine su uno stato discriminativo su qualcosa nel mondo. La semantica di tali affermazioni rende chiaro cosa si crede che siano i colori: proprietà rifrangenti delle superfici degli oggetti, o di volumi trasparenti (il cubetto di ghiaccio rosa, il fascio di luce del proiettore). E questo è proprio quello che sono infatti – sebbene dire esattamente *quali* proprietà rifrangenti siano non è facile (per ragioni che indagheremo nel prossimo paragrafo).

Ma i nostri stati discriminativi interni non hanno *anche* qualche proprietà «intrinseca» speciale, le proprietà soggettive, private, ineffabili che costituiscono *il modo in cui le cose ci appaiono* (in cui le vediamo, le sentiamo, le udiamo, ecc.)? Queste proprietà addizionali sarebbero i qualia, e prima di esaminare gli argomenti che i filosofi hanno escogitato nel tentativo di *dimostrare* che queste proprietà addizionali esistono, proveremo a rimuovere in primo luogo i motivi per credere a queste proprietà, trovando spiegazioni alternative dei fenomeni che sembrano richiederle. Allora le pecche sistematiche di quei tentativi dimostrativi saranno immediatamente visibili.

Secondo questa prospettiva alternativa, i colori *sono* proprietà «lì fuori», dopo tutto. Al posto della «idea di rosso» lockiana abbiamo (negli esseri umani normali) stati discriminativi che hanno il contenuto: *rosso*. Un esempio ci aiuterà a chiarire in maniera perfetta cosa sono questi stati discriminativi – e più importante, cosa non sono. Noi possiamo paragonare i colori delle cose nel mondo mettendole fianco a fianco e guardandole, lasciando così sorgere un giudizio in noi; ma possiamo anche paragonare il colore delle cose semplicemente ricordandole o immaginandole «nella nostra mente». Il rosso standard delle strisce della bandiera americana è lo stesso rosso, o è più scuro o più chiaro o più luminoso o dà più o meno sull'arancione del rosso standard del vestito di Babbo Natale (o di una buca per le lettere inglese o della stella rossa sovietica)? (Se non riesci a ricordare nessuna di queste coppie, tenta con un'altra, per esempio il blu del cielo e il blu dei caschi delle truppe dell'ONU, o il verde del feltro del tavolo da biliardo e il verde della bandiera italiana, il giallo del limone e il giallo del burro.) Noi siamo capaci di fare questi confronti «con l'occhio della mente», e quando li facciamo, in qualche modo facciamo avvenire qualcosa in noi che richiama le informazioni dalla memoria e ci permette di paragonare, nell'esperienza cosciente, i colori degli oggetti standard come li ricordiamo (come abbiamo l'abitudine di ricordarli, in ogni caso). Alcuni di noi ci riescono meglio di altri, senza dubbio, e molti di noi non si fidano troppo dei giudizi che otteniamo in queste circostanze. Ecco perché ci portiamo a casa dei campioni di vernici o di tessuti, così possiamo mettere uno accanto all'altro due esempi dei colori che vogliamo paragonare.

Quando facciamo questi paragoni «con l'occhio della mente», cosa accade secondo il mio punto di vista? Qualcosa di strettamente analogo a ciò che accade in una macchina – un robot – che sia in grado di fare un paragone simile. Si ripensi al *Vorsetzer* CADBLIND Mark I del capitolo 10 (quello con la telecamera che poteva inquadrare lo schermo CAD). Supponiamo di mettergli di fronte una fotografia a colori di Babbo Natale e di chiedergli se il rosso della fotografia è più intenso del rosso della bandiera americana (che ha già immagazzinato nella sua memoria). Ecco cosa farebbe: richiamerebbe la sua rappresentazione della «Old Glory» dalla memoria e individuerebbe le strisce «rosse» (che nel suo diagramma sono etichettate «rosso 163»). Comparerebbe poi questo rosso con il rosso del vestito di Babbo Natale nella fotografia di fronte alla sua

telecamera, che sarebbe trasdotto dal suo sistema grafico a colori come, poniamo, rosso 172. Comparerebbe i due rossi *sottraendo 163 da 172 e ottenendo 9*, che interpreterebbe, diciamo, come la prova che il rosso di Babbo Natale sembra un po' più intenso e profondo (a *lui*) del rosso della bandiera americana.

Questa descrizione è volutamente molto semplificata, per rendere più incisiva l'asserzione che voglio fare: è ovvio che il CADBLIND Mark I non usa un figmento per rendere la rappresentazione tratta dalla memoria (o dalla sua percezione attuale), *ma neanche noi lo usiamo*. Il sistema CADBLIND Mark I ha – lo ammetto – uno spazio cromatico molto semplice e povero, con ben poche delle associazioni o inclinazioni strutturali connaturate nello spazio cromatico degli esseri umani, ma, a parte questa enorme differenza nella complessità disposizionale, non c'è una differenza importante. Potrei dire perfino che non c'è una differenza *qualitativa* tra l'esecuzione di un tale compito da parte del CADBLIND e la nostra. Gli stati discriminativi del CADBLIND Mark I possiedono un contenuto esattamente nello stesso modo degli stati discriminativi cerebrali che io ho messo al posto delle idee di Locke, ed esattamente per le stesse ragioni. Il CADBLIND Mark I *sicuramente* non ha alcun qualia (almeno, questo è quello che mi aspetto che gli amanti dei qualia si affretterebbero a dire ora), quindi segue necessariamente dal mio paragone che io sto sostenendo che anche noi non abbiamo qualia. Ciò che sto fermamente negando è il *tipo* di differenza che la gente immagina ci sia tra un soggetto di esperienza umano e uno meccanico (si ricordi la macchina assaggiatrice di vino che abbiamo immaginato nel capitolo 2): non c'è alcuna differenza del genere. Sembra esserci soltanto.

2. *Perché ci sono i colori?*

Quando Otto, nel capitolo 11, giudicò che sembrava esserci un lucente anello rosa, qual era il contenuto del suo giudizio? Se, come ho detto, il suo giudizio non era su un qualia, una proprietà di un sembrar-anello «fenomenico» (fatto di figmento), su che diamine era? Quale proprietà egli si scoprì tentato di attribuire (falsamente) a qualcosa lì fuori nel mondo?

Molti hanno notato che è curiosamente difficile dire esattamente quali proprietà delle cose nel mondo possano essere i colori. L'idea più semplice e attraente – che si può ancora

trovare in molte trattazioni elementari – è che ogni colore può essere associato con un'unica lunghezza d'onda della luce, e quindi che la proprietà dell'essere rosso è semplicemente la proprietà di riflettere tutte le lunghezze d'onda rosse della luce e di assorbire tutte le altre lunghezze d'onda. Ma si sa, non da ieri, che questo è falso. Superfici con differenti proprietà riflettenti fondamentali possono essere viste come lo stesso colore e la stessa superficie in differenti condizioni d'illuminazione può essere vista come colori differenti. Le lunghezze d'onda della luce che entra nell'occhio sono correlate solo indirettamente ai colori che vediamo negli oggetti. (Vedi Gouras, 1984; Hilbert, 1987; e Hardin, 1988 per delle rassegne dei dettagli con diverse accentuazioni.) Per quelli che hanno sperato che ci fosse un modo semplice ed elegante di incassare la cambiale di Locke sui poteri disposizionali delle superfici, la situazione non potrebbe essere più deprimente. Alcuni (per es., Hilbert, 1987) hanno deciso di ancorare oggettivamente il colore, dichiarando che è una proprietà relativamente diretta degli oggetti esterni, come la proprietà di «riflettenza spettrale della superficie»; avendo fatto questa scelta, devono poi concludere che la visione cromatica normale spesso ci presenta delle illusioni, poiché le costanti che percepiamo si accoppiano così malamente alle costanti della riflettenza spettrale della superficie misurate dagli strumenti scientifici. Altri hanno concluso che è meglio considerare soggettivamente le proprietà del colore, come proprietà che devono essere definite strettamente in termini di sistemi di stati cerebrali negli osservatori, ignorando la ingannevole variabilità presente nel mondo che fa sorgere questi stati: «Gli oggetti colorati sono illusioni, ma non illusioni infondate. Noi siamo normalmente in stati percettivi cromatici, e questi sono stati neurali» (Hardin, 1988, p. 111; vedi Thompson, Palacios e Varela, in stampa, per una discussione critica su queste opzioni e per ulteriori argomenti a favore della migliore opzione da adottare).

Ciò che è indubitabile è che non esiste una proprietà semplice e non disgiuntiva delle superfici tale che tutte e soltanto le superfici con quella proprietà siano rosse (nel senso delle qualità secondarie di Locke). Questo è un fatto a prima vista enigmatico, e perfino deprimente, poiché sembra suggerire che la nostra presa percettiva sul mondo sia molto peggiore di quanto avessimo pensato – che stiamo vivendo in una specie di mondo sognato o che siamo vittime di un'illusione di massa. La nostra visione cromatica non ci dà accesso a proprietà semplici

degli oggetti, anche se sembra darcela. Perché dovrebbe essere così?

Solo sfortuna? Cattiva progettazione? Niente affatto. C'è una prospettiva differente, e molto più illuminante, che possiamo adottare sul colore, e che mi è stata mostrata per la prima volta dal filosofo della neuroscienza Kathleen Akins (1989, 1990).[1] Talvolta nuove proprietà giungono all'esistenza per una ragione. Un esempio particolarmente utile viene offerto dal famoso caso di Julius e Ethel Rosenberg, che nel 1953 furono accusati di spionaggio e condannati a morte per aver consegnato all'Unione Sovietica dei segreti sul progetto di bomba atomica statunitense. Ad un certo punto del processo venne fuori che avevano improvvisato un astuto sistema di identificazione: una scatola di cartone di una confezione di budino era stata strappata in due e i pezzi erano consegnati a due persone che dovevano fare molta attenzione nell'identificarsi tra loro. Ogni pezzo lacerato divenne un «rilevatore» del proprio compagno praticamente unico e a prova d'errore: ad un incontro successivo ogni parte avrebbe potuto produrre il suo pezzo ed essere sicuro di aver incontrato la persona giusta solo se i due pezzi avessero combaciato perfettamente. Perché questo sistema funziona? Perché strappare in due pezzi il cartone produce un bordo di una tale complessità informazionale che sarebbe virtualmente impossibile riprodurlo deliberatamente. (Si noti che tagliando la scatola di cartone con riga e lametta non si otterrebbe certo lo stesso scopo.) Il particolare bordo frastagliato di un pezzo diventa un congegno di riconoscimento della configurazione del suo compagno *praticamente* unico; è un apparato o un trasduttore per rilevare la proprietà formale *M*, dove *M* è posseduta esclusivamente dal suo compagno.

In altre parole, la proprietà formale *M* e il rilevatore della proprietà *M* che la rileva sono fatti l'uno per l'altro. Non ci sarebbe alcuna ragione per entrambi di esistere, di essere stati creati, in assenza dell'altro. E lo stesso si può dire per i colori e la visione cromatica: sono stati fatti l'uno per l'altro. La codificazione tramite colore è un'idea abbastanza recente nella «ingegneria dei fattori umani», ma i suoi pregi sono ora ampiamente riconosciuti. Nei corridoi degli ospedali vengono tracciate delle linee colorate per semplificare le indicazioni che i pazienti devono seguire: «Per andare in fisioterapia, devi solo seguire la linea gialla; per andare al centro trasfusionale, segui la linea rossa!». I costruttori di televisori, calcolatori e altre apparecchiature elettroniche identificano con i colori le enormi matasse

di fili che si trovano all'interno di questi apparecchi di modo che è più facile seguirli da un punto all'altro. Queste sono solo applicazioni recenti, ma naturalmente l'idea è molto più vecchia; più vecchia della Lettera Scarlatta con cui si marcavano le adultere, più vecchia delle uniformi colorate usate per distinguere gli amici dai nemici nel cuore della battaglia, più vecchia della specie umana, in realtà.

Noi siamo inclini a pensare alla codificazione tramite colore come a una introduzione intelligente di schemi colorati «convenzionali» progettati per sfruttare l'esistenza della «naturale» visione cromatica, ma così si perde di vista il fatto che la «naturale» visione cromatica *fin dall'inizio* è evoluta assieme ai colori la cui *raison d'être* era la codificazione tramite colore (Humphrey, 1976). Alcune cose nella natura «avevano bisogno di essere viste» e altre avevano bisogno di vederle, così si è evoluto un sistema che tendeva a minimizzare il compito per le ultime rafforzando la rilevanza delle prime. Si considerino gli insetti. La loro visione cromatica è evoluta assieme ai colori delle piante che impollinavano, un buon trucco progettuale di cui hanno beneficiato entrambi. Senza la codificazione tramite colore dei fiori, la visione cromatica degli insetti non si sarebbe evoluta, e viceversa. Così il principio della codificazione tramite colore sta alla base della visione cromatica degli insetti, non è solo un'invenzione recente di una specie intelligente di mammiferi. Lo stesso si può dire dell'evoluzione della visione cromatica in altre specie. Anche se alcuni generi di visione cromatica possono essere evoluti inizialmente per poter discriminare visivamente dei fenomeni inorganici, non è ancora chiaro se questo sia realmente avvenuto in qualche specie di questo pianeta. (Evan Thompson mi ha fatto notare che le api possono usare il loro tipo speciale di visione cromatica durante il volo, per discriminare la luce solare polarizzata nei giorni nuvolosi, ma non è questa un'utilizzazione secondaria della visione cromatica che originalmente si è evoluta assieme ai colori dei fiori?)

Differenti sistemi di visione cromatica sono evoluti indipendentemente, talvolta con spazi colorati radicalmente differenti. (Per una breve rassegna, e riferimenti, vedi Thompson, Palacios e Varela, in stampa.) Non tutte le creature con occhi hanno una qualche visione cromatica. Gli uccelli, i pesci, i rettili e gli insetti hanno sicuramente una visione cromatica, più o meno come il nostro sistema «tricromatico» (rosso-verde-blu); cani e gatti non ce l'hanno. Tra i mammiferi, solo i primati hanno una visione cromatica e tra loro ci sono sorprendenti differenze.

Quali specie hanno la visione cromatica, e perché? Questa si rivela una questione complessa e affascinante, ancora piena di semplici congetture.

Perché le mele diventano rosse quando maturano? È naturale presupporre che sia possibile dare una risposta completa chiamando in causa i cambiamenti chimici che si verificano quando lo zucchero ed altre sostanze raggiungono determinate concentrazioni nel frutto in via di maturazione, il che produce varie reazioni, e così via. Ma così si ignora il fatto che tanto per cominciare non ci sarebbero mele se non ci fossero «mangiatori di mele sparpagliatori di semi» che le vedessero, quindi il fatto che le mele siano facilmente visibili ad almeno qualche varietà di mangiatori di mele è una condizione della loro esistenza, non un semplice «caso» (dal punto di vista delle mele!). Il fatto che le mele abbiano le loro specifiche proprietà di riflettenza spettrale superficiale dipende tanto dai fotopigmenti che erano disponibili per essere imbrigliati nelle cellule coni negli occhi dei fruttivori, quanto dagli effetti delle interazioni tra lo zucchero e gli altri componenti nella chimica del frutto. La frutta che non è codificata tramite colore viene messa in difficoltà dalla concorrenza sui banchi del supermercato della natura, ma la pubblicità falsa verrà punita; la frutta che è matura (piena di nutrimento) e *che pubblicizza questo fatto* venderà meglio, ma la pubblicità deve essere ritagliata sulle capacità e le inclinazioni visive di un consumatore specifico.

In principio, i colori erano fatti per essere visti da quelli che erano fatti per vederli. Ma questo connubio si è evoluto gradualmente, a caso, sfruttando qualunque vantaggio imprevisto veniva apportato dai materiali disponibili, rielaborando infinite volte le opportunità di un nuovo Trucco, e tollerando sempre una grande quantità di inutili variazioni e inutili (puramente coincidenti) costanti. Queste costanti coincidenti riguardavano spesso caratteristiche «più fondamentali» del mondo fisico. Una volta comparse delle creature che potevano distinguere le bacche rosse da quelle verdi, esse potevano distinguere anche i rubini rossi dagli smeraldi verdi, ma questo regalo era dovuto solo ad una coincidenza. Il fatto che ci sia una differenza *di colore* tra i rubini e gli smeraldi può allora essere considerato un fenomeno cromatico *derivato*. Perché il cielo è blu? Perché le mele sono rosse e l'uva viola, non il contrario.

È un errore pensare che prima ci fossero i colori – le rocce colorate, l'acqua colorata, il cielo colorato, la ruggine rossastra e il cobalto blu brillante – e poi sia intervenuta Madre Natura

che ha sfruttato *quelle* proprietà usandole per codificare le cose tramite colori. Piuttosto, prima c'erano varie proprietà riflettenti delle superfici, varie proprietà reattive dei fotopigmenti, e così via, e poi Madre Natura partendo da questi materiali grezzi ha sviluppato coppie efficienti e mutuamente adattate di sistemi di visione dei «colori» e codificazioni di «colori», e tra le proprietà che sono sortite da questo processo progettuale ci sono le proprietà che noi esseri umani normali chiamiamo colori. Se accade che il blu del cobalto e il blu delle ali di una farfalla combaciano (nella visione degli esseri umani normali) ciò è solo una coincidenza, un trascurabile effetto collaterale dei processi che hanno portato all'esistenza la visione cromatica e che perciò (come Locke stesso potrebbe sottoscrivere) hanno dato un nome a un insieme curiosamente assemblato di complessi di proprietà primarie con la proprietà secondaria comune di produrre lo stesso effetto in un insieme di osservatori normali.

«Ma comunque», si vorrà obiettare, «prima ancora che ci fossero animali con visione cromatica, c'erano meravigliosi tramonti rossi e brillanti smeraldi verdi!» Beh, sì, si può dire così, ma allora quegli stessi tramonti sarebbero stati anche accecanti, multicolori e disgustosi, resi in colori che non possiamo vedere e quindi nemmeno nominare. Si deve, cioè, ammettere questo se esistono o *potrebbero esistere* su qualche pianeta creature con apparati sensori che sarebbero così influenzati da tali smeraldi e tramonti. E per quanto ne sappiamo, ci potrebbe essere, da qualche parte, una specie che vede naturalmente due (o diciassette) colori differenti in un gruppo di smeraldi che *noi* troviamo indistinguibilmente verdi.

Molti esseri umani sono incapaci di distinguere tra il verde e il rosso. Supponiamo che lo fossimo tutti: sarebbe allora una convinzione comune che sia i rubini sia gli smeraldi siano «rordi» – dopo tutto, a osservatori normali sembrano proprio come le altre cose rorde: gli autocarri dei pompieri, i prati ben innaffiati, le mele mature e acerbe (Dennett, 1969). Se arrivasse della gente come noi e insistesse a dire che i rubini e gli smeraldi sono in effetti di due colori differenti, non ci sarebbe modo di dichiarare uno di questi sistemi di visione cromatica più «vero» dell'altro.

Il filosofo Jonathan Bennett (1965) attira la nostra attenzione su una circostanza che, più convincentemente, mette in luce la stessa posizione, in un'altra modalità sensoriale. La sostanza fenoltiourea, ci dice, ha un gusto amaro per un quarto

della popolazione mondiale ed è completamente priva di gusto per il resto. Il gusto che una persona prova è determinato geneticamente. La fenoltiourea è amara o insapore? Tramite «eugenetica» o ingegneria genetica, potremmo riuscire ad eliminare il genotipo responsabile del gusto amaro del fenolo. Se ci riuscissimo, la fenoltiourea sarebbe allora *paradigmaticamente* insapore, così insapore come l'acqua distillata: insapore per tutti gli esseri umani. Se realizzassimo l'esperimento genetico opposto, renderemmo allora la fenoltiourea paradigmaticamente amara. Ora, prima che ci fosse qualunque essere umano, la fenoltiourea era *sia* amara *sia* insapore? Era chimicamente identica ad ora.

I fatti sulle qualità secondarie sono inevitabilmente legati a una classe di riferimento di osservatori, ma ci sono modi forti o deboli di trattare il legame. Noi possiamo dire che le qualità secondarie sono *adorabili* piuttosto che *sospette*. Una ragazza potrebbe essere adorabile anche se non fosse mai accaduto che sia stata vista da un qualche osservatore che la trovi adorabile, ma non potrebbe – logicamente – essere sospetta fino a che qualcuno non la sospetti davvero di qualcosa. Si può dire che dei casi particolari di qualità adorabili (come la qualità dell'adorabilità) esistano come disposizioni lockiane prima del momento (eventuale) in cui esercitano il loro potere su un osservatore, producendo in lui gli effetti che le definiscono. Allora qualche donna non vista (cresciuta da sola su un'isola, immagino) potrebbe essere autenticamente adorabile, avendo i poteri disposizionali di influenzare gli osservatori normali di una certa classe in un certo modo, malgrado non abbia mai avuto l'opportunità di farlo. Ma le qualità adorabili non possono essere definite indipendentemente dalle inclinazioni, dalle suscettibilità, dalle disposizioni di una classe di osservatori, così non ha alcun senso parlare dell'esistenza di proprietà adorabili indipendentemente dall'esistenza di osservatori specifici. In realtà, questo è un po' troppo forte. Le qualità adorabili non *sarebbero* definite – non avrebbe senso definir*le*, a differenza di tutte le altre proprietà che è logicamente possibile assemblare e manipolare – indipendentemente da una tale classe di osservatori. Così mentre sarebbe logicamente possibile («retrospettivamente», si potrebbe dire) raccogliere insieme le cose che hanno la stessa proprietà cromatica tramite qualcosa di simile alla enumerazione per forza bruta, le ragioni per scegliere tali proprietà (per esempio, allo scopo di spiegare certe regolarità causali in un insieme di oggetti curiosamente complicato) dipende dall'esistenza della classe di osservatori.

Gli elefanti marini sono adorabili? Non per noi. È difficile immaginare una creatura più brutta. Ciò che rende un elefante marino adorabile per un altro elefante marino non è ciò che rende una donna adorabile per un uomo, e chiamare adorabile una donna che non è mai stata osservata e che potrebbe attrarre fortemente un elefante marino equivarrebbe ad abusare sia di lei che del termine. È soltanto in riferimento ai gusti umani, che sono caratteristiche contingenti e particolarmente specifiche del mondo, che la proprietà dell'adorabilità (per gli esseri umani) può essere identificata.

D'altra parte, le qualità sospette (come la qualità di essere una persona sospetta) vengono intese in una maniera tale che presuppone che qualunque cosa che possieda la proprietà abbia già provocato l'effetto che la definisce in almeno un osservatore. Qualcuno può meritarsi il più profondo sospetto – può perfino essere ovviamente colpevole – ma non può essere una persona sospetta fino a che qualcuno non lo abbia effettivamente sospettato. Io non sto sostenendo che i colori siano qualità sospette. La nostra convinzione intuitiva che uno smeraldo non ancora visto in un blocco di minerale sia *già* verde non deve essere negata. Io sto sostenendo, invece, che i colori sono qualità adorabili, la cui esistenza, legata com'è a una classe di riferimento di osservatori, non ha senso in un mondo in cui gli osservatori non abbiano posto. Questo è più facile da accettare per alcune qualità secondarie che per altre. Che i vapori di zolfo sputati fuori dai vulcani primordiali fossero gialli sembra in un certo senso più oggettivo del dire che puzzassero, ma fintanto che ciò che intendiamo con «giallo» è ciò che *noi* intendiamo con giallo, le due affermazioni sono parallele. Supponiamo infatti che un terremoto primordiale avesse sollevato il versante di una collina mettendo a nudo le strisce formate dalle centinaia di falde chimicamente differenti. Queste strisce sarebbero *visibili*? Dobbiamo chiedere da chi? Forse noi potremmo vedere alcune di esse ed altre no. Forse alcune delle strisce invisibili sarebbero visibili per un piccione tetracromatico o per creature che vedano nella porzione infrarossa o ultravioletta dello spettro elettromagnetico. Per la stessa ragione non ha senso chiedersi se la differenza tra smeraldi e rubini sia una differenza visibile senza specificare il sistema visivo in questione.

L'evoluzione ammorbidisce il «soggettivismo» o «relativismo» implicito nel fatto che le qualità secondarie siano qualità adorabili. Mostra che l'assenza di caratteri comuni «semplici» o «fondamentali» nelle cose che sono tutte dello stesso colore non

è un segno di illusione totale, ma piuttosto di una tolleranza diffusa per le rilevazioni «positive false» delle proprietà ecologiche che contano davvero.[2] Le categorie basilari del nostro spazio cromatico (e naturalmente del nostro spazio odoroso e del nostro spazio sonoro, e di tutto il resto) sono forgiate da pressioni selettive, cosicché in generale ha senso chiedersi a che serve una particolare discriminazione o preferenza. Ci sono delle ragioni del perché evitiamo gli odori di determinate cose e ne cerchiamo altri, perché preferiamo certi colori ad altri, perché certi suoni ci danno più fastidio o ci calmano di più. Non sempre sono le *nostre* ragioni, ma piuttosto le ragioni di antenati lontani, che lasciano le loro tracce fossili nelle predisposizioni incorporate che forgiano naturalmente i nostri spazi qualitativi. Ma da buoni darwiniani, dovremmo anche riconoscere la possibilità – in realtà, la necessità – di altre predisposizioni non funzionali, distribuite a caso nella popolazione dalla variazione genetica. Affinché la pressione selettiva possa favorire in modo differenziale quelli che mostrano una predisposizione contro *F* una volta che *F* divenga ecologicamente importante, deve esserci stata una variazione inutile (non ancora funzionale) nella «attitudine verso *F*» su cui la selezione possa agire. Per esempio, se mangiare trippa *dovesse* implicare un nefasto destino preriproduttivo, solo quelli tra noi che fossero «naturalmente» (e *prima d'ora* inutilmente) disposti contro la trippa avrebbero un vantaggio (forse minuscolo all'inizio, ma ben presto esplosivo, se le condizioni lo favoriscono). Così dal fatto che trovi qualcosa (per es., i broccoli) indescrivibilmente e ineffabilmente sgradevoli non segue che c'è una ragione per ciò. E non segue neppure che sei difettoso se non sei d'accordo con i tuoi simili su ciò. Potrebbe semplicemente trattarsi di una delle innate propensioni nel tuo spazio qualitativo che non ha, o non ancora per lo meno, nessuna significanza funzionale. (E per il tuo bene, devi sperare che se mai avesse una significanza, è perché i broccoli si sono improvvisamente rivelati nocivi per noi.)

Queste considerazioni evoluzionistiche ci aiutano molto a capire perché le qualità secondarie si rivelano così «ineffabili», così refrattarie alla definizione. Come la proprietà formale *M* del pezzo di cartone dei Rosenberg, le qualità secondarie sono estremamente difficili da definire direttamente. Fa parte dell'essenza del trucco dei Rosenberg il fatto che non possiamo sostituire il nostro posticcio predicato *M* con una descrizione della proprietà più complessa, ma più accurata ed esaustiva, perché, se potessimo farlo, potremmo (noi o qualcun altro) usare

quella descrizione come una ricetta per produrre un altro caso di *M* o un altro rilevatore di *M*. I nostri rilevatori di qualità secondarie non furono progettati specificamente per rilevare solo proprietà difficili da definire, ma il risultato non cambia di molto. Come Akins (1989) osserva, i nostri sistemi sensoriali non sono *fatti per* rilevare proprietà «basilari» o «naturali» dell'ambiente, ma solo per servire i nostri scopi «narcisistici» nel mantenerci in vita; la natura non costruisce macchine epistemiche.

L'unico modo *facilmente disponibile* per dire quale proprietà formale sia *M* è di indicare il rilevatore di *M* e dire che *M* è la proprietà formale rilevata da questa cosa qui. Nella stessa difficoltà si trova naturalmente chi prova a dire quale proprietà sta rilevando (o rilevando erroneamente) quando qualcosa «sembra nel modo in cui gli sembra». Così ora possiamo rispondere alla domanda con cui abbiamo aperto questo paragrafo: quale proprietà Otto giudica che una cosa abbia quando egli giudica che è rosa? La proprietà che egli chiama rosa. E che proprietà è questa? È difficile dirlo, ma questo non dovrebbe imbarazzarci, perché possiamo dire perché è difficile dirlo. La cosa migliore che possiamo praticamente fare quando ci chiedono quali proprietà superficiali noi rileviamo con la visione cromatica, è di dire, in modo non molto informativo, che rileviamo le proprietà che rileviamo. Se qualcuno vuole una trattazione più informativa su queste proprietà, può consultare una letteratura piuttosto vasta e incomprensibile di biologia, neuroscienza e psicofisica. E Otto non può dire niente di più sulla proprietà che egli chiama rosa dicendo «*È così*!» (indicando al suo «interno» una proprietà privata e fenomenica della sua esperienza). Questa mossa non fa altro che indicare il proprio stato personale di discriminazione del colore, una mossa parallela a quella di tenere in mano un pezzo di cartone e dire che esso rileva *questa* proprietà formale. Otto indica il suo congegno discriminativo, forse, ma non un qualche qualia che trasudi da esso o venga indossato da esso o reso da esso, quando fa il suo lavoro. Non esistono cose del genere.

Ma tuttavia [insiste Otto], non hai ancora detto perché il rosa dovrebbe sembrare *così*!

Così come?

Così. Come la «rosità» particolarmente ineffabile, meravigliosa,

intrinseca di cui sto godendo in questo momento. *Questa* non è qualche proprietà riflettente di superfici indescrivibilmente convolute degli oggetti esterni.

Otto, vedo che stai usando il termine *godendo*. Non sei il solo. Spesso, quando un autore vuole sottolineare che l'argomento è passato dalla (mera) neuroanatomia all'esperienza, dalla (mera) psicofisica alla coscienza, dalla (mera) informazione ai qualia, spinge sul palcoscenico la parola «godere».

3. *Godiamoci le nostre esperienze*

> Ma Dan, i qualia sono ciò che rende la vita degna di essere vissuta!
>
> WILFRID SELLARS (accanto ad una bottiglia di eccellente Chambertin, Cincinnati, 1971)

> Se ciò che voglio quando bevo dell'ottimo vino è l'informazione sulle sue proprietà chimiche, perché non leggo semplicemente l'etichetta?
>
> SYDNEY SHOEMAKER, Tufts Colloquium, 1988

Alcuni colori furono fatti per piacere, e lo stesso è vero per alcuni odori e alcuni gusti. Altri colori, odori e gusti furono fatti per dispiacere. Per esprimere lo stesso concetto un po' più precisamente, non è un caso che noi (e altre creature che possono rilevarli) amiamo e odiamo colori, odori, gusti e altre qualità secondarie. Così come abbiamo ereditato evoluti rilevatori di simmetria verticale nei nostri sistemi visivi che ci avvertono (noi come i nostri antenati) del fatto ecologicamente significante che un'altra creatura ci sta guardando, così abbiamo ereditato evoluti rilevatori di qualità che non sono degli informatori disinteressati, ma piuttosto dei segnalatori e sentinelle: sirene, sia nel senso dei pompieri sia nel senso omerico.

Come abbiamo visto nel capitolo 7, sull'evoluzione, questi allarmisti originari sono stati successivamente cooptati in una schiera di organizzazioni più complicate, fatte di milioni di associazioni, e forgiati, nel caso dell'uomo, da migliaia di memi. In questo modo l'attrazione bruta – che vuole essere subito soddisfatta – del sesso e del cibo, e l'avversione bruta – che fa scappare immediatamente – del dolore e della paura vengono mescolate insieme in un'infinità di combinazioni. Quando un

organismo scopre che è pagante attendere di fronte a qualche caratteristica del mondo *malgrado* la sua avversione naturale a farlo, deve costruire qualche coalizione per controbilanciare l'avversione e impedirle di vincere. La tensione semi-stabile che ne risulta può poi diventare essa stessa un gusto acquisito, da ricercare in certe condizioni. Quando un organismo scopre che deve soffocare gli effetti di certi segnalatori insistenti se vuole mantenere la propria condotta, può coltivare un gusto per qualunque sequenza di attività a sua disposizione che tenda a produrre la pace e la quiete desiderate. In tal modo potremmo giungere ad amare il cibo piccante che brucia la nostra bocca (Rozin, 1982), la musica deliziosamente «discordante», e sia il tranquillo e calmo realismo di Andrew Wyeth che lo sconvolgente e ardente espressionismo di Willem de Kooning. Marshall McLuhan (1967) ha proclamato che il mezzo è il messaggio, una quasi verità che è più vera forse nel sistema nervoso che in qualunque altro forum di comunicazione. Ciò che vogliamo quando sorseggiamo un buon vino non è, infatti, l'informazione sui suoi contenuti chimici; ciò che vogliamo è *di essere informati* sui suoi contenuti chimici nel nostro modo preferito. E le nostre preferenze sono *in definitiva* basate sulle predisposizioni ancora presenti nei nostri sistemi nervosi malgrado la loro significanza ecologica possa essere scomparsa milioni di anni fa.

La nostra stessa tecnologia ci ha tenuto in larga parte nascosto questo fatto. Come lo psicologo Nicholas Humphrey nota,

> non appena mi guardo attorno nella stanza in cui lavoro, innumerevoli colori prodotti dall'uomo mi assalgono da ogni superficie: libri, cuscini, un tappeto sul pavimento, una tazzina da caffè, una scatola di punti metallici – blu, rossi, gialli, verdi di notevole intensità. Ci sono altrettanti colori qui che nella foresta tropicale. Eppure, mentre quasi ogni colore nella foresta avrebbe un significato, qui nel mio studio quasi nulla lo ha. L'anarchia cromatica ha preso il sopravvento. (1983, p. 149)

Si consideri, per esempio, il fatto curioso che le scimmie non amano la luce rossa. Se possono scegliere, le scimmie rhesus mostrano una spiccata preferenza per l'estremità blu-verde dello spettro e diventano agitate quando sono obbligate a stare in ambienti rossi (Humphrey, 1972, 1973, 1983; Humphrey e Keeble, 1978). Perché è così? Humphrey dice che il rosso è sempre usato per l'allarme, il colore estremo della codificazione

tramite colore, ma per questa stessa ragione è ambiguo: il frutto rosso può essere buono da mangiare, ma il serpente o l'insetto rosso sta probabilmente segnalando che è velenoso. Quindi il «rosso» invia messaggi contrastanti. Ma tanto per cominciare come mai invia un messaggio di «allarme»? Forse perché è il contrasto disponibile più forte con l'ambiente di sfondo del verde vegetativo o del blu marino, o – nel caso delle scimmie – perché la luce rossa (da rossa a arancione-rossastra a arancione) è la luce dell'alba e del tramonto, il momento del giorno in cui quasi tutti i predatori delle scimmie sono a caccia.

Le proprietà affettive o emotive del rosso non si limitano alle scimmie rhesus. Tutti i primati condividono queste reazioni, inclusi gli esseri umani. Se i lavoratori della tua fabbrica si attardano nei servizi igienici, una bella mano di rosso nei gabinetti risolverà il problema – ma ne creerà degli altri (vedi Humphrey, di prossima pubblicazione). Tali reazioni «viscerali» non sono ristrette ai colori, naturalmente. La reazione che la maggior parte dei primati cresciuti in cattività hanno quando vedono per la prima volta un serpente mette chiaramente in luce che li aborriscono profondamente, ed è probabile che il disgusto umano tradizionale per i serpenti abbia un'origine biologica che spiega quella biblica, piuttosto che l'inverso.[3] Cioè, la nostra eredità biologica favorisce i memi dell'odio verso i serpenti.

Ecco ora due differenti spiegazioni del disagio che la maggior parte di noi sente (anche se «riesce a dominarlo») quando vediamo un serpente:

1. I serpenti evocano in noi un particolare e intrinseco quale di schifo-serpentino quando li vediamo, e il nostro disagio è una reazione a questo quale.
2. Ci scopriamo tutt'altro che impazienti di vedere i serpenti a causa di predisposizioni innate incorporate nei nostri sistemi nervosi. Queste favoriscono il rilascio di adrenalina, danno l'avvio a sottoprogrammi di combattimento o di fuga e, attivando vari legami associativi, presentano vari scenari che implicano pericolo, violenza e danneggiamenti. L'originaria avversione dei primati viene, in noi, modificata, revisionata, deviata in centinaia di modi dai memi che l'hanno sfruttata, cooptata, forgiata. (Noi potremmo formulare una spiegazione di tipo «funzionalista» come questa a vari livelli differenti. Per esempio, potremmo permetterci di parlare più liberamente sulle capacità delle percezioni di serpenti di

produrre ansietà, paure, anticipazioni del dolore, e simili, ma ciò potrebbe essere visto come «barare» e così lo evito.)

Il guaio del primo tipo di spiegazione è che ha solo l'apparenza di una spiegazione. L'idea che una proprietà «intrinseca» (di rosa eveniente, di schifo-serpentino, di dolore, di aroma del caffè) possa *spiegare* le reazioni di un soggetto a una circostanza è senza speranze – un chiaro caso di *virtus dormitiva* (vedi p. 77). Dichiarare una teoria colpevole di contenere una vacua *virtus dormitiva* non è così semplice, però. Talvolta è perfettamente sensato postulare una *virtus dormitiva* temporanea, in attesa di ulteriori indagini. Il concepimento è, per definizione potremmo dire, la causa della gravidanza. Se non avessimo altri modi di identificare il concepimento, dire ad una donna che è incinta perché ha concepito sarebbe una mossa inutile, non una spiegazione. Ma, una volta conosciuta la necessaria teoria meccanica del concepimento, possiamo capire *come* il concepimento sia la causa della gravidanza, e la capacità informativa viene ristabilita. Nello stesso spirito, potremmo identificare i qualia, per definizione, come la causa più prossima del nostro godimento e della nostra sofferenza (dicendolo in modo grossolano), e poi procedere ad assolvere i nostri obblighi informativi intraprendendo il secondo stile di spiegazione. Ma, piuttosto curiosamente, i qualiofili (come io chiamo quelli che ancora credono nei qualia) non la pensano così; essi insistono, come Otto, che i qualia «ridotti» a meri complessi di disposizioni a reagire realizzati meccanicamente non sono i qualia di cui loro stanno parlando. I *loro* qualia sono qualcosa di differente.

Considera [dice Otto] il modo in cui l'anello rosa mi sembra proprio *ora*, in questo preciso istante, separato da tutte le mie disposizioni, le associazioni passate e le attività future. *Questo*, il *modo* purificato, isolato *in cui mi trovo rispetto* al colore in questo momento – questo è il mio quale rosa.

Otto ha appena fatto un errore. Infatti, questo è un grande errore, la fonte di tutti i paradossi sui qualia, come vedremo. Ma prima di esporre la follia insita nel prendere questa strada, voglio mostrare alcuni dei vantaggi della strada scartata da Otto: la strada «riduzionista» di *identificare* «il modo in cui mi trovo rispetto a qualcosa» con la somma totale di tutte le disposizioni reattive personali inerenti al mio sistema nervoso

che risultano dalla presenza di una certa configurazione di stimolazione.

Si consideri che cosa si dovesse provare a essere un abitante luterano di Lipsia che, diciamo, nel 1725 andava in chiesa a sentire una delle cantate corali di J. S. Bach nella sua prima esecuzione. (Questo esercizio di immaginare *che cosa si prova* è una preparazione per il capitolo 14, dove ci interesseremo della coscienza negli altri animali.) Probabilmente non ci sono significative differenze biologiche tra noi oggi e i luterani tedeschi del diciottesimo secolo; siamo della stessa specie e il tempo che è passato è infinitesimale. Ma, a causa della tremenda influenza della cultura – la memosfera – il nostro mondo psicologico è piuttosto differente dal loro, in maniere che avrebbero un notevole impatto sulle nostre rispettive esperienze al primo ascolto di una cantata di Bach. La nostra immaginazione musicale è stata arricchita e complicata in mille modi (da Mozart, da Charlie Parker, dai Beatles), ma ha perso anche alcune potenti associazioni sulle quali Bach poteva contare. Le sue cantate corali sono state costruite attorno a corali e melodie di inni tradizionali che erano profondamente familiari ai fedeli dell'epoca e che quindi provocavano ondate di associazioni emotive e tematiche non appena le loro tracce e echi fossero percepite nella musica. Molti di noi oggi conoscono questi corali solo attraverso l'arrangiamento di Bach, così quando li ascoltiamo, li sentiamo con un orecchio differente. Se vogliamo immaginare che cosa si provava ad essere un ascoltatore lipsiano di Bach, non basta ascoltare le stesse note sugli stessi strumenti nello stesso ordine; dobbiamo anche prepararci in qualche modo a rispondere a quelle note con gli stessi batticuori, fremiti e onde di nostalgia.

Non è completamente impossibile preparare se stessi in questo modo. Un musicologo che avesse evitato con cura ogni contatto con la musica posteriore al 1725 e avesse acquisito una profonda familiarità con la musica tradizionale di quel periodo sarebbe una buona prima approssimazione. La cosa più importante, come mostrano queste osservazioni, è che non è impossibile sapere esattamente in quali modi dovremmo preparare noi stessi, indipendentemente dall'avere o meno l'effettiva intenzione di sobbarcarci tutto il lavorio necessario. Così potremmo sapere che cosa si prova «in astratto», per così dire, e di fatto lo ho appena detto: gli abitanti di Lipsia, ascoltando le cantate corali, erano portati a ricordare tutte quelle associazioni che davano un sapore particolare al loro riconoscimento delle

melodie corali. È abbastanza facile immaginare cosa provavano – sebbene ciò possa avvenire solo grazie a variazioni tratte dalla nostra esperienza personale. Possiamo immaginare che cosa si proverebbe a sentire l'arrangiamento di Bach di familiari canti di Natale, per esempio, o di *Quel mazzolin di fiori*. Non possiamo realizzare il compito esattamente, ma solo perché non possiamo dimenticare o abbandonare tutto quello che noi sappiamo e che gli abitanti di Lipsia non sapevano.

Per rendersi conto di quanto sia cruciale questo nostro eccesso di bagaglio, si immagini che i musicologi scoprano una cantata di Bach fino ad ora sconosciuta, sicuramente di quel grand'uomo, ma rimasta sempre in un cassetto e mai ascoltata da nessuno finora, nemmeno dal suo compositore. Tutti sarebbero ansiosi di ascoltarla, di sperimentare per la prima volta i «qualia» che gli abitanti di Lipsia avrebbero conosciuto, se solo l'avessero sentita; ma ciò si rivela impossibile, perché il tema principale della cantata, per una sfortunata coincidenza, è identico alle prime cinque note della canzone dei sette nani *Andiam, andiam a lavorar*! Noi che siamo gravati da quel motivo non saremo *mai* in grado di ascoltare la versione di Bach come egli la intendeva o come gli abitanti di Lipsia l'avrebbero recepita.

Un caso più chiaro di blocco dell'immaginazione sarebbe difficile da trovare, ma si noti che ciò non ha nulla a che fare con differenze biologiche o anche con proprietà «intrinseche» o «ineffabili» della musica di Bach. La ragione per la quale non possiamo rivivere con l'immaginazione nei minimi dettagli l'esperienza musicale degli abitanti di Lipsia è semplicemente che dovremmo portare noi stessi in questo viaggio immaginario, e noi sappiamo troppo. Ma se vogliamo, possiamo elencare con precisione le differenze tra le nostre disposizioni e conoscenze e le loro, e, comparando le due liste, giungere a valutare, *con il dettaglio che vogliamo*, le differenze tra cosa provavano loro ad ascoltare Bach e cosa proviamo noi. Sebbene possiamo lamentarci di questa inaccessibilità, almeno possiamo capirla. Non ci sarebbe nessun mistero residuo; solo un'esperienza che potrebbe essere descritta piuttosto accuratamente, ma non direttamente goduta a meno che non ci impegnassimo nel lunghissimo e ridicolo travaglio di ricostruzione delle nostre strutture disposizionali personali.

I qualiofili, comunque, hanno resistito a questa conclusione. A loro è sembrato che sebbene un'indagine come quella appena immaginata potrebbe risolvere *quasi* tutti gli interrogativi che ci

ponevamo su cosa provassero gli abitanti di Lipsia, ci sarebbe sempre un residuo ineffabile, *qualcosa* riguardo a cosa sentivano i Lipsiani che nessun progresso ulteriore nelle nostre conoscenze «disposizionali» o «meccaniche» potrebbe ridurre a zero. Ecco perché i qualia devono essere chiamati in causa dai qualiofili come caratteristiche *addizionali*, che oltrepassano e sono separate dal cablaggio che determina l'indietreggiamento, l'aggrottamento delle ciglia, le urla e gli altri «semplici comportamenti» del disgusto, ribrezzo e paura. Possiamo vedere questo con chiarezza se torniamo al nostro esempio dei colori.

Supponiamo di suggerire ad Otto che ciò che faceva del suo «rosa eveniente» l'esperienza particolarmente allettante di cui stava godendo era semplicemente la somma totale di tutte le associazioni e disposizioni reattive innate e apprese che sono state scatenate dal modo particolare in cui egli è stato informato (erroneamente) dai suoi occhi:

I qualia, Otto, *sono solo* quei complessi di disposizioni. Quando tu dici «*Questo* è il mio quale», ciò che tu stai individuando, o a cui ti stai riferendo, *che tu lo capisca o no*, è solo il tuo complesso strettamente personale di disposizioni. *Sembra* che tu ti riferisca a un qualcosa di privato e ineffabile nel tuo occhio della mente, una tonalità privata di rosa omogeneo, ma così è solo come sembra a te, non come è. Quel tuo «quale» è un personaggio verosimile nel mondo fittizio della tua eterofenomenologia, ma ciò che si rivela essere nel mondo *reale* del tuo cervello è solo un complesso di disposizioni.

Questo non può essere tutto ciò che c'è in esso [replica Otto, compiendo il passo fatale nella tradizione qualiofila], giacché sebbene quel complesso di mere disposizioni potrebbe essere la base o la fonte, in qualche modo, del mio particolare quale di rosa, tutte queste disposizioni potrebbero essere cambiate senza cambiare il mio intrinseco quale, oppure il mio intrinseco quale potrebbe cambiare, senza cambiare quell'aggregato di mere disposizioni. Per esempio, i miei qualia potrebbero essere *invertiti* senza invertire tutte le mie disposizioni. Potrei avere tutte le reattività e le associazioni che ora ho per il *verde* in congiunzione al *qualia* che ora ho per il *rosso*, e viceversa.

L'idea della possibilità di una tale «inversione dei qualia» è uno dei memi filosofici più virulenti. Locke l'ha discussa nel suo *Saggio sull'intelletto umano* (1690), e molti miei studenti mi hanno confessato di essersi imbattuti da soli nella stessa idea da bambini e di esserne rimasti affascinati. L'idea sembra essere del tutto chiara e innocua:

C'è la maniera in cui le cose appaiono a me, e risuonano per me, e odorano per me, e così via. Tutto ciò è ovvio. Mi chiedo, però, se la maniera in cui le cose appaiono a me è la stessa maniera in cui appaiono agli altri.

I filosofi hanno composto molte variazioni differenti su questo tema, ma la versione classica è quella interpersonale: come posso sapere che tu ed io vediamo lo stesso colore *soggettivo* quando guardiamo qualcosa? Poiché entrambi abbiamo appreso le parole che si riferiscono ai colori quando ci hanno mostrato, in un mondo pubblico, vari oggetti colorati, i nostri comportamenti verbali combaceranno anche se nelle nostre esperienze ci saranno dei colori soggettivi completamente differenti – anche se la maniera in cui le cose rosse appaiono a me è la maniera in cui le cose verdi appaiono a te, per esempio. Chiameremmo le stesse cose pubbliche «rosso» e «verde» anche se le nostre esperienze private fossero «opposte» (o solo differenti).

Esiste un modo di sapere se le cose stanno davvero così? Consideriamo l'ipotesi che le cose rosse appaiono uguali a te e a me. Questa ipotesi è sia irrefutabile che non confermabile? Molti hanno pensato che effettivamente è cosi, ed alcuni hanno concluso che proprio per la stessa ragione è una sorta di non senso, malgrado il suo iniziale appello al senso comune. Altri si sono domandati se sia possibile immaginare che la tecnologia venga in nostro aiuto e confermi (o refuti) l'ipotesi dell'inversione interpersonale dello spettro. Il film di fantascienza *Brainstorm* (che non è, mi affretto a dire, una versione cinematografica del mio libro *Brainstorms*) mette in evidenza proprio l'adatto congegno immaginario: un particolare apparato neuroscientifico si adatta alla tua testa e via cavo alimenta il mio cervello con la tua esperienza visiva. Con gli occhi chiusi riferisco accuratamente tutto ciò che stai guardando, ma mi meraviglio di come il cielo sia giallo, l'erba rossa e così via. Se

avessimo una macchina del genere, un tale esperimento non confermerebbe, empiricamente, l'ipotesi che i nostri qualia sono differenti? Ma supponiamo che un tecnico sfili la spina del cavo di connessione, la giri di 180 gradi e la reinserisca nella presa; ora io riferisco che il cielo è blu, l'erba verde e così via. Quale sarebbe la posizione «giusta» della spina? La progettazione e la costruzione di un tale congegno – supponendo per un momento che sia possibile – richiederebbe che la sua «fedeltà» sia accordata o calibrata tramite la normalizzazione dei resoconti dei due soggetti, e così saremmo risospinti al nostro punto di partenza. Ora si potrebbe cercare di allontanare questa conclusione specificando ulteriori dettagli, ma è opinione diffusa tra i qualiofili che questa è una causa persa; sembra esserci un accordo generale che la morale di *questo* esperimento mentale sia che non sarebbe possibile alcuna comparazione intersoggettiva di qualia, neppure con una tecnologia perfetta. Tutto ciò gioca, però, a favore del disgustoso punto di vista «verificazionista» o «positivista» che l'idea stessa dell'inversione dei qualia è un non senso – e quindi l'idea stessa di qualia è un non senso. Come disse il filosofo Ludwig Wittgenstein, usando la famosa analogia del «coleottero nella scatola»,

> la cosa contenuta nella scatola non fa parte in nessun caso del gioco linguistico; nemmeno come un *qualcosa*: infatti la scatola potrebbe anche essere vuota. – No, si può «dividere per» la cosa che è nella scatola; di qualunque cosa si tratta si annulla. (1953, I, 293)

Ma che cosa significa ciò esattamente? Significa che i qualia sono reali ma privi di effetti? O che non ci sono affatto qualia, dopo tutto? È tuttavia sembrato ovvio a molti filosofi che hanno riflettuto su ciò che i qualia siano reali, anche se una differenza tra i qualia sarebbe una differenza impossibile da rilevare. Così stavano le cose, purtroppo, fino a che qualcuno non ha immaginato la versione presumibilmente migliore dell'esperimento di pensiero: l'inversione *intra*personale dello spettro. Sembra che l'idea sia venuta in mente indipendentemente a varie persone (Gert, 1965; Putnam, 1965; Taylor, 1966: Shoemaker, 1969; Lycan, 1973). In questa versione, le esperienze da comparare sono tutte in un'unica mente, così non abbiamo bisogno della inutilizzabile macchina di *Brainstorm*.

Una mattina ti svegli e scopri che l'erba è diventata rossa, il cielo giallo e così via. Nessun altro nota alcuna anomalia cromatica

nel mondo e quindi il problema deve essere in te. Sei autorizzato a concludere, sembra, di aver subito un'inversione dei qualia della visione cromatica. Come è successo? Si scopre in seguito che, mentre dormivi, alcuni neurochirurghi malvagi hanno invertito tutti i fili – i neuroni – che provengono dalle cellule coniche sensibili ai colori nelle tue retine.

Fin qui, tutto bene. L'effetto su di te sarebbe impressionante, forse perfino terrificante. Tu saresti certamente capace di rilevare che ora le cose ti appaiono in una maniera molto differente da prima, e avremmo anche una spiegazione scientifica appropriata di questo fatto: il grappolo di neuroni nella corteccia visiva che «si occupa dei» colori, per esempio, riceverebbe i propri stimoli da un insieme sistematicamente spostato di recettori retinici. Sembra che così si sia vinta metà della battaglia: una differenza nei qualia sarebbe rilevabile dopo tutto, se fosse una differenza che si sviluppasse piuttosto rapidamente in un'unica persona.[4] Ma questa è solo metà della battaglia, giacché l'immaginaria burla neurochirurgica ha invertito anche tutte le tue disposizioni reattive; non solo tu *dici* che la tua esperienza cromatica è stata del tutto scombussolata, ma anche il tuo comportamento non verbale relativo ai colori è stato invertito. Il nervosismo che eri solito mostrare alla luce rossa lo mostri ora alla luce verde, e hai perso la spigliatezza con cui ti affidavi normalmente nella tua vita a vari schemi codificati dai colori (se giochi a pallacanestro con i Boston Celtics, seguiti a passare erroneamente la palla ai giocatori in maglia rossa*).

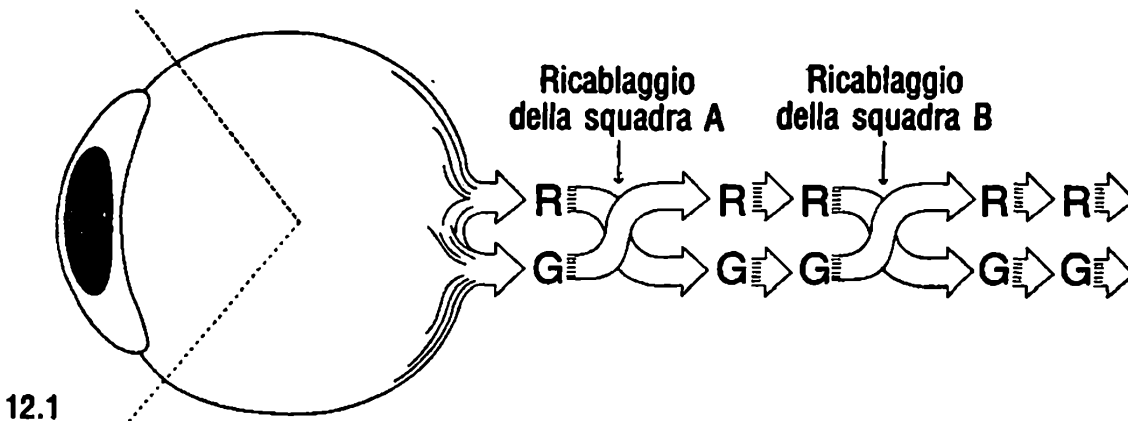
Ciò di cui il qualiofilo ha bisogno è un esperimento di pensiero che dimostri che la maniera in cui le cose appaiono possa essere indipendente da tutte queste disposizioni reattive. Così possiamo complicare la storia con uno sviluppo ulteriore; dobbiamo immaginare che avvenga qualcosa che disfa l'inversione delle disposizioni reattive mentre lascia intatti i «qualia» invertiti.

Qui la letteratura si imbarca in fantasie sempre più involute, perché nessuno pensa nemmeno per un momento che la maniera in cui le cose appaiono sia mai *davvero* separata dalle disposizioni reattive del soggetto; solo che questa è ritenuta una importante *possibilità in linea di principio* dai qualiofili. Per mostrare ciò hanno bisogno di descrivere un caso possibile, per quanto bizzarro, in

* La maglia dei Boston Celtics è verde (N.d.T.).

cui sarebbe ovvio che questo distacco fosse reale. Si consideri una storia che *non* funzionerebbe:

Una notte mentre dormi, alcuni neurochirurghi malvagi invertono tutti i fili che provengono dai coni (proprio come prima), e poi, poco dopo nella stessa notte movimentata, un'altra squadra di neurochirurghi, la squadra B, interviene ed effettua un ricablaggio *complementare* un po' più su nel nervo ottico.

12.1

Ciò ristabilisce tutte le vecchie disposizioni reattive (possiamo presumere), ma, purtroppo, ristabilisce anche i vecchi qualia. Alle cellule nella corteccia che «si occupano del» colore, per esempio, ora arriveranno di nuovo i segnali originari, grazie alla rapida riparazione del danno effettuata dalla squadra B. Il secondo capovolgimento sembra avvenuto troppo presto; è avvenuto lungo la strada *verso* l'esperienza cosciente. Così dobbiamo modificare la nostra storia: il secondo capovolgimento avviene più tardi, *dopo* che i qualia invertiti abbiano fatto il loro inchino di presentazione nella coscienza, ma *prima* che qualunque reazione invertita ad essi possa essere avviata. Ma è possibile ciò? No, se gli argomenti a favore del modello delle Molteplici Versioni sono corretti. Nella «catena» causale che va dal globo oculare alla coscienza al comportamento successivo non può essere tracciata alcuna linea tale che tutte le reazioni a x avvengano dopo di essa e la coscienza di x avvenga prima di essa. Questo perché non c'è una semplice catena causale, ma piuttosto una rete causale, con tragitti multipli su cui vengono edite simultaneamente e semi-indipendentemente Molteplici Versioni. La descrizione qualiofila avrebbe un senso se ci fosse un Teatro Cartesiano, un luogo speciale nel cervello dove avviene l'esperienza cosciente. Se ci fosse un luogo simile, lo potremmo inserire tra i due capovolgimenti, lasciando i qualia

invertiti nel Teatro, mantenendo normalizzate tutte le disposizioni reattive. Poiché non esiste un simile Teatro Cartesiano, comunque, l'esperimento di pensiero non ha senso. Non c'è un modo coerente per descrivere la situazione necessaria. Non c'è modo di separare le proprietà *presentate* nella coscienza dalle molteplici reazioni del cervello alle sue discriminazioni, perché non c'è un tale processo di presentazione addizionale.

Nella letteratura sullo spettro invertito, si suppone spesso che il secondo capovolgimento sia realizzato non tramite chirurgia ma grazie al graduale adattamento del soggetto al nuovo regime di esperienze. Questo sembra a prima vista sensato; la gente riesce ad adattarsi incredibilmente bene a bizzarre modificazioni sensoriali. Sono stati svolti molti esperimenti sull'inversione del campo visivo in cui i soggetti indossavano occhialoni che rovesciavano ogni cosa – mettendo sottosopra l'immagine retinica! (Per es., Stratton, 1896; Kohler, 1961; Welch, 1978, fornisce un buon riassunto; vedi anche Cole, 1990). Dopo aver indossato costantemente per vari giorni gli occhialoni invertenti di un tipo o dell'altro (ce ne sono vari, alcuni hanno un campo visivo ampio, altri danno come una visione attraverso un tunnel), i soggetti spesso mostrano un adattamento sorprendentemente positivo. Nel film di Ivo Kohler sui suoi esperimenti a Innsbruck, vediamo due suoi soggetti – comicamente goffi quando indossano gli occhiali per la prima volta – che sciano lungo ripidi pendii e vanno in bicicletta nel traffico cittadino, indossando sempre gli occhialoni invertenti, e sembrando completamente adattati ad essi.

Supponiamo quindi che ti sia gradualmente adattato all'inversione chirurgica della tua visione cromatica. (Perché mai vorresti adattarti, o dovresti adattarti, è un'altra faccenda, ma possiamo concedere anche questo ai qualiofili, non farà che accelerare la loro sconfitta.) Ora qualche adattamento sarebbe dapprincipio chiaramente post-*esperienziale*. Possiamo supporre che il cielo sereno ti *apparirebbe* ancora giallo, ma inizieresti a chiamarlo blu per tenerti al passo con i tuoi vicini. La vista di un oggetto sconosciuto potrebbe provocarti una momentanea confusione: «È ver– cioè *rosso*!». E che dire del tuo nervosismo alla luce verde – apparirebbe ancora come un'anormalità nella tua risposta galvanica cutanea? Ai fini della sua argomentazione, il qualiofilo deve immaginare, per quanto inverosimile possa essere, che *tutte* le tue disposizioni reattive si adattino, lasciando indietro solo il residuo dei qualia ancora invertiti; così a beneficio dell'argomento, concediamo che anche le più fonda-

mentali e le più innate predisposizioni nel tuo spazio qualitativo si «adattino» – questo è incredibile, ma il peggio deve ancora venire.

Per raffigurare il quadro necessario, il qualiofilo deve supporre che tutti questi adattamenti alla fine divengano una seconda natura: rapidi e non calcolati. (Se non fosse così ci sarebbero delle disposizioni reattive residue che sarebbero ancora differenti, e l'argomento richiede che tutte le differenze siano bandite.) Quindi lasciamo che sia così. Ora, presupponendo che *tutte* le tue disposizioni reattive siano ristabilite, qual è la tua idea intuitiva sui tuoi qualia? Sono ancora invertiti o no?

A questo punto sarebbe legittimo non pronunciarsi, sulla base del fatto che dopo aver tollerato tante assunzioni dubbie a beneficio dell'argomento, o ti ritrovi completamente vuoto – nessuna intuizione viene in superficie – o scopri di non fidarti di nessuna intuizione che ti colpisca. Ma forse ti sembrerebbe piuttosto ovvio che i tuoi qualia fossero ancora invertiti. Ma perché? Che cosa nella descrizione appena fatta ti ha condotto a pensarla così? Forse, anche se hai seguito tutte le istruzioni, hai aggiunto senza volerlo qualche assunzione non richiesta o non sei riuscito ad intravedere alcune possibilità non escluse dalla situazione. Io suggerisco che la spiegazione più verosimile della tua convinzione intuitiva che, in questa situazione immaginaria, avresti ancora i «qualia invertiti» è che stai facendo l'assunzione aggiuntiva e ingiustificata che tutta l'adattazione avvenga nel «lato post-esperienziale».

Non potrebbe essere, però, che l'adattamento venga realizzato nel tragitto verso la coscienza? Quando si indossano per la prima volta degli occhialoni molto scuri, non si vedrà alcun colore – o perlomeno i colori che si vedono sono strani e difficilmente riconoscibili – ma dopo averli portati per un po' ritorna una visione cromatica sorprendentemente normale. (Cole, 1990, attira l'attenzione dei filosofi su questi effetti, che se vuoi puoi sperimentare da te con gli occhiali all'infrarosso per la visione notturna che puoi trovare in svendite di residuati militari.) Forse, non conoscendo questo fatto sorprendente, non ti è mai venuto in mente che *potresti* adattarti all'intervento chirurgico in un modo molto simile. Avremmo potuto attirare l'attenzione su questa possibilità nell'esperimento di pensiero, aggiungendo qualche dettaglio:

... E man mano che l'adattamento procede, ti stupisci spesso

che i colori delle cose non ti sembrano più così strani dopo tutto, e talvolta ti confondi e ti capita di fare delle *doppie* correzioni: quando ti si chiede il colore di un oggetto mai visto, dici «È ver–, no rosso – no, *è* verde!»

Raccontando la storia in questa maniera, potrebbe sembrare «ovvio» che i qualia cromatici stessi si siano adattati, o siano stati reinvertiti. Ma in ogni caso, potresti pensare ora, le cose devono stare in un modo o nell'altro. Non è possibile che non sia sempre perfettamente ovvio che tipo di aggiustamento sia stato realizzato! La presupposizione non esaminata su cui si basa *questa* convinzione è che tutti gli adattamenti possano essere categorizzati o come pre-esperienziali o come post-esperienziali (staliniani o orwelliani). Dapprincipio questa può sembrare una presupposizione innocua, poiché i casi estremi sono facili da classificare. Quando il cervello compensa i movimenti degli occhi e della testa, producendo un mondo visivo stabile «nell'esperienza», ciò è sicuramente una cancellazione pre-esperienziale di effetti, un adattamento sul tragitto ascendente verso la coscienza. E quando immagini di fare delle compensazioni periferiche («tardive») nella scelta delle parole riferite ai colori («È ver-, cioè *rosso*!») ciò è ovviamente un aggiustamento meramente comportamentale, post-esperienziale. Non è allora conforme a ragione che quando *tutti* gli adattamenti siano stati fatti, o lasciano invertito il colore soggettivo (il colore «nella coscienza») oppure no? Ecco come potremmo dirlo: aggiungi vari capovolgimenti nel tragitto in ascesa; se sono in numero pari – come nel lavoro ben fatto della squadra B – i qualia sono normalizzati, se sono dispari, i qualia sono ancora invertiti. Assolutamente insensato. Ricordati della curva neo lafferiana nel capitolo 5. Non è affatto una necessità logica o geometrica che ci sia un unico valore di una variabile discriminata che possa essere scelta come il valore della variabile «nella coscienza».

Possiamo dimostrare questo con l'aiuto di una piccola fantasia di nostra produzione, giocata con le regole dei qualiofili. Supponiamo che prima dell'intervento chirurgico una certa tonalità di blu tendesse a farti ricordare un'automobile contro la quale molto tempo fa sei andato a sbattere, e quindi era un colore che evitavi. Subito dopo l'intervento chirurgico non dimostri nessuna reazione negativa nei confronti delle cose di quel colore, trovandole, possiamo supporre, di un giallo innocuo e privo di memorie. Dopo esserti completamente adattato,

però, tu eviti di nuovo le cose con quella tonalità di blu, *e questo perché ti ricordano di quell'incidente* (se non te lo ricordassero, ci sarebbe una disposizione reattiva non adattata). Ma se ti chiedessimo se ciò avviene perché, ricordando l'incidente, l'automobile era gialla – proprio come quel fastidioso oggetto davanti a te ora – o perché, ricordando l'incidente, l'automobile era blu – proprio come quel fastidioso oggetto davanti a te ora – tu non dovresti realmente essere in grado di rispondere. Il tuo comportamento verbale sarà completamente «adattato»; alla domanda «Di che colore era l'automobile contro cui ha sbattuto?», la tua risposta immediata, come seconda natura sarà «blu», e tu chiamerai blu senza esitare anche l'oggetto fastidioso di fronte a te. Questo implica che hai *dimenticato* il lungo periodo di allenamento?

No. Non abbiamo bisogno di qualcosa di così drammatico come l'amnesia per spiegare la tua incapacità a rispondere, giacché abbiamo un'abbondanza di casi quotidiani in cui sorge lo stesso fenomeno. Ti piace la birra? Molte persone a cui piace la birra saranno d'accordo che la birra è un gusto acquisito. Ci si allena gradualmente – o semplicemente si arriva – ad amare quel sapore. Quale sapore? Il sapore del primo sorso?

A nessuno potrebbe piacere *quel* sapore [un bevitore di birra di lunga data potrebbe ribattere]. Per un bevitore di lunga data la birra ha un gusto differente. Se la birra avesse seguitato ad avere il gusto del primo sorso, non sarei certo diventato un bevitore di birra! O, per esprimere lo stesso concetto all'inverso, se il primo sorso di birra avesse avuto il gusto che ha ora, non avrei certo avuto bisogno di abituarmi a quel gusto! Mi sarebbe piaciuto il primo sorso proprio come quello che ho appena bevuto.

Se questo bevitore di birra ha ragione, allora la birra *non* è un gusto acquisito. Nessuno arriva ad apprezzare *il gusto del primo sorso*. Il gusto della birra, invece, *cambia gradualmente* per il bevitore. Altri bevitori di birra potrebbero insistere che, no, la birra ha ora *esattamente lo stesso* gusto che aveva la prima volta, solo che ora apprezzano *proprio quel gusto*. C'è una reale differenza? C'è una differenza nell'eterofenomenologia, certamente, e la differenza deve essere spiegata. *Potrebbe essere* che le differenti convinzioni sgorghino da genuine differenze nella capacità discriminativa del seguente genere: nel primo tipo di bevitore di birra «l'allenamento» ha modificato «la forma» dello spazio

qualitativo del gusto, mentre nel secondo tipo lo spazio qualitativo rimane approssimativamente identico, ma «la funzione valutativa» su quello spazio è stata revisionata. Oppure *potrebbe essere* che alcuni o perfino tutti i bevitori di birra si stanno prendendo in giro da soli (come quelli che insistono nel dire che le Marilyne ad alta risoluzione *sono* realmente tutte nello sfondo del loro campo visivo). Dobbiamo andare al di là dei mondi eterofenomenologici e osservare gli effettivi eventi cerebrali per capire se c'è un'interpretazione (sia pur «stiracchiata») che possa conservare la verità delle pretese dei bevitori di birra, e, *se* c'è, sarà soltanto perché decidiamo di *ridurre* «il gusto» ad un qualche complesso di disposizioni reattive (Dennett, 1988a). Dovremmo «distruggere» i qualia per poterli «salvare».

Così se un bevitore di birra aggrotta le sopracciglia, assume un'espressione estremamente seria, e dice che ciò a cui si sta riferendo è «*il gusto che ha ora la birra per me*», sta sicuramente ingannandosi se pensa *con ciò* di potersi riferire ad un qualia di sua conoscenza, uno stato soggettivo che è indipendente dalle sue mutevoli attitudini reattive. Gli sembra di poterlo fare, ma non può.[5]

E per lo stesso ragionamento, nel caso immaginario dell'incidente automobilistico che ritorna alla mente a causa dell'oggetto blu, ti inganneresti da solo se pensassi di poter stabilire, *dal modo in cui l'oggetto ti appare*, se era «intrinsecamente» lo stesso modo in cui l'automobile ti appariva quando sei andato a sbattere. Ciò basta a tagliare le gambe all'esperimento di pensiero del qualiofilo, perché lo scopo era di descrivere un caso in cui fosse *ovvio* che i qualia sarebbero invertiti mentre le disposizioni reattive sarebbero normalizzate. La *presupposizione* che si possa stabilire la differenza è ingiustificata, e senza la presupposizione non c'è argomentazione ma solo una pompa di intuizione – una storia che ti spinge a dichiarare le tue intuizioni viscerali senza dartene una buona ragione.

Ingiustificata o no, può *ancora* sembrare del tutto ovvio che «i colori soggettivi che vedresti nelle cose» *dovrebbero* essere «in un modo o nell'altro». Questo mostra soltanto la potente forza gravitazionale che il Teatro Cartesiano esercita sulle nostre immaginazioni. Per aiutarci a far crollare le attrattive residue di questa idea, consideriamo ulteriormente il parallelo con gli occhiali che invertono le immagini. Una volta che l'adattamento di questi soggetti che indossano gli occhialoni è diventato talmente una seconda natura che possono sciare e andare in bicicletta, la domanda naturale (ma fuorviante) da fare è

questa: si sono adattati *rovesciando di nuovo nella posizione giusta il loro mondo visivo o abituandosi al loro mondo visivo rovesciato*? E cosa dicono loro? Dicono cose differenti, in correlazione abbastanza esplicita con il grado del loro adattamento. Più l'adattamento è completo, più i soggetti liquidano la domanda come impropria o indecidibile. Il che è esattamente ciò che la teoria delle Molteplici Versioni richiede: poiché ci sono una miriade di discriminazioni e reazioni che devono essere messe a punto, diffuse su tutto il cervello, alcune di esse che trattano con «riflessi» di basso livello (come abbassarsi nella maniera giusta quando qualcosa si avvicina rapidamente) e altre che trattano con azioni deliberate che richiedono un'attenzione focalizzata, non sorprende che man mano che gli adattamenti in questo groviglio si accumulano, i soggetti perdano ogni convinzione di dire «le cose sembrano nella stessa maniera di prima» invece di dire «le cose sembrano ancora differenti, ma mi ci sono abituato». In un certo senso le cose sembrano come prima a loro (a giudicare dalle loro reazioni), in altri sensi sembrano differenti (a giudicare da altre reazioni). Se ci fosse un'unica rappresentazione dello spazio visivo-motorio attraverso cui tutte le reazioni agli stimoli visivi dovessero essere incanalate, dovrebbe essere «in un modo o nell'altro», forse, ma non c'è un'unica rappresentazione del genere. *Il* modo in cui le cose appaiono a loro è composto di molte abitudini reattive parzialmente indipendenti, non da un'unica immagine nella testa intrinsecamente dritta o rovesciata. L'unica cosa importante è l'accordo tra l'input e l'output, e poiché ciò viene realizzato in molti posti differenti con molti mezzi differenti e molto indipendenti, non si può dire cosa «conta» come «il mio campo visivo è ancora rovesciato».

Lo stesso si può dire per l'inversione dei «qualia». L'idea che essa sia qualcosa *in aggiunta* all'inversione di tutte le proprie disposizioni reattive, così che se queste ultime fossero rinormalizzate l'inversione dei qualia rimarrebbe, è semplicemente parte del mito tenace del Teatro Cartesiano. Questo mito è celebrato nell'elaborato esperimento di pensiero sull'inversione dello spettro, ma celebrare non significa corroborare o dimostrare. *Se* non ci sono qualia oltre la somma totale delle disposizioni a reagire, l'idea di tenere i qualia costanti mentre si modificano le disposizioni è autocontraddittoria.

5. *I qualia sono «epifenomenici»?*

C'è un altro esperimento di pensiero filosofico sulla nostra esperienza del colore che si è dimostrato irresistibile; è quello presentato da Frank Jackson (1982): il caso molto discusso di Mary, la scienziata del colore che non ha mai visto i colori. Come un buon esperimento di pensiero, mette immediatamente in evidenza il suo punto di forza anche per il non iniziato; ma in realtà è un cattivo esperimento di pensiero, una pompa di intuizioni che ci incoraggia a fraintendere le premesse!

> Mary è una scienziata brillante che, per qualche ragione, è costretta a indagare il mondo da una stanza in bianco e nero *attraverso* uno schermo televisivo in bianco e nero. Ella si specializza in neurofisiologia della visione e acquisisce tutte le informazioni fisiche ottenibili su ciò che avviene in noi quando vediamo i pomodori maturi, o il cielo, e usa termini come *rosso*, *blu* e così via. Per esempio, scopre esattamente quali combinazioni di lunghezze d'onda dal cielo stimolano la retina, e esattamente come ciò produce, *via* il sistema nervoso centrale, la contrazione delle corde vocali e l'espulsione dell'aria dai polmoni che risulta nella pronuncia della frase «il cielo è blu»... Che cosa succede quando a Mary viene permesso di lasciare la sua stanza in bianco e nero o le viene regalato uno schermo a colori? *Apprenderà* qualcosa o no? Sembra del tutto ovvio che apprenderà qualcosa sul mondo e su come esso appare nella nostra esperienza visiva. Ma allora è inevitabile dire che la sua conoscenza precedente era incompleta. Eppure aveva *tutte* le informazioni fisiche. *Ergo* c'è qualcos'altro che bisogna avere, e il fisicalismo è falso [...] (p. 128)

È difficile che l'argomento possa essere presentato in maniera più chiara. Mary *non ha mai avuto alcuna* esperienza del colore (nella sua stanza non ci sono specchi in cui guardarsi la faccia, è obbligata a indossare guanti neri, ecc., ecc.), e così, nel momento fatidico in cui i suoi carcerieri decidono infine di lasciarla entrare in quel mondo colorato che lei conosce solo per descrizione (e immagini in bianco e nero), «sembra del tutto ovvio», come dice Jackson, che imparerà qualcosa. Infatti, possiamo tutti immaginarcela vividamente mentre vede per la prima volta una rosa rossa ed esclama «Dunque, *questo* è il rosso!». E possiamo anche pensare che se le prime cose colorate che le vengono mostrate sono dei blocchi di legno colorato, e le viene detto che uno di essi è rosso, ella non avrà la più pallida idea di quale esso sia fino a che non imparerà in qualche modo come il linguaggio si accoppia alle sue nuove esperienze.

Questo è il modo in cui quasi tutti immaginano questo esperimento di pensiero, non solo i non addetti ai lavori, ma anche i filosofi più perspicaci e accorti (Tye, 1986; Lewis, 1988; Loar, 1990; Lycan, 1990; Nemirov, 1990; Harman, 1990; Block, 1990; van Gulick, 1990). Solo Paul Churchland (1985, 1990) ha offerto una qualche seria resistenza all'*immagine* della drammatica scoperta di Mary, così come viene vividamente evocata dall'esperimento di pensiero. L'immagine è errata; se questo è il modo in cui immagini il caso, semplicemente non stai seguendo le indicazioni! La ragione per cui nessuno segue le indicazioni è che ciò che viene chiesto di immaginare è così ridicolmente immenso che nemmeno ci si può provare. La premessa cruciale è che «lei ha *tutte* le informazioni fisiche». Questo non è facilmente immaginabile, così nessuno se ne preoccupa. Si immagina semplicemente che lei sappia moltissimo – forse si immagina che lei conosca tutto quello che *oggi* si sa sulla neurofisiologia della visione cromatica. Ma questo è solo una goccia nel mare, e non è sorprendente che Mary imparerebbe qualcosa se *questo* è tutto quello che sapeva.

Per mettere in luce l'inganno dell'immaginazione, mi si lasci continuare la storia in un modo imprevisto – ma legittimo:

E così, un giorno, i carcerieri di Mary decisero che era giunto il momento per lei di vedere i colori. Per la sua prima esperienza cromatica, prepararono uno scherzo: pitturarono di un bel blu brillante una banana matura. Appena Mary la vide, disse: «Ehi! Cercavate di prendermi in giro! Le banane sono gialle, ma questa è blu!». I suoi carcerieri rimasero stupefatti. Come ha fatto? «Semplice», replicò lei, «dovete ricordarvi che io so *ogni cosa* – assolutamente ogni cosa – si possa mai sapere sulle cause e gli effetti fisici della visione cromatica. Così prima che mi faceste vedere la banana, naturalmente io sapevo già, in tutti i minimi dettagli, quali impressioni fisiche esatte un oggetto giallo o un oggetto blu (o uno verde, ecc.) avrebbe provocato sul mio sistema nervoso. Così sapevo già quali *pensieri* avrei avuto (perché, dopo tutto, la "mera disposizione" a pensare su questo o quello non è uno dei vostri famosi qualia?). Io non sono stata minimamente sorpresa della mia esperienza del blu (ciò che mi ha stupito è che abbiate tentato uno scherzo così a buon mercato ai miei danni). Capisco che sia *difficile per voi immaginare* che io possa sapere così tanto sulle mie disposizioni reattive che l'effetto prodotto in me dal blu non mi sorprenda. Naturalmente è difficile per voi da immaginare. È difficile per chiunque immaginare le conseguenze del fatto che qualcuno abbia assolutamente tutte le informazioni fisiche su qualunque cosa!»

Sicuramente ho barato, starai pensando. Devo aver nascosto qualche impossibilità dietro il velo delle parole di Mary. Puoi dimostrarlo? Io non sto sostenendo che il mio epilogo della storia dimostra che Mary *non* impari qualcosa, sostengo soltanto che il modo usuale di immaginare la storia *non* dimostra che lo faccia. Non dimostra nulla; pompa semplicemente le intuizioni a favore della possibilità che impari («sembra così ovvio») facendoti cullare nell'illusione di aver immaginato effettivamente ciò che le premesse richiedono.

È naturalmente vero che in ogni versione realistica, facilmente immaginabile della storia, Mary giungerebbe ad imparare qualcosa, ma in ogni versione realistica, facilmente immaginabile, ella potrebbe sapere molto, ma non avrebbe tutte le informazioni fisiche su qualcosa. Immaginare semplicemente che Mary conosca molto, e fermarsi qui, non è un buon modo per capire le implicazioni dell'ipotesi che lei abbia «tutte le informazioni fisiche» – proprio come immaginarsi che ella sia schifosamente ricca non sarebbe una buona maniera per immaginarsi l'implicazione dell'ipotesi che possegga ogni cosa. Possiamo riuscire ad immaginare meglio l'estensione dei poteri che la sua conoscenza le conferisce, se iniziamo ad enumerare alcune delle cose che già sa. Conosce il bianco e il nero e le tonalità di grigio, conosce le differenze tra il colore di ogni oggetto e tali proprietà superficiali come la lucidità opposta alla opacità, e sa tutto sulla differenza tra i confini di luminanza e i confini cromatici (i confini di luminanza sono quelli che emergono su un televisore in bianco e nero, in parole povere). E conosce precisamente quali effetti – descritti in termini neurofisiologici – ogni colore particolare avrà sul suo sistema nervoso. Così l'unico compito che le rimane è di immaginare il modo per identificare quegli effetti neurofisiologici «dall'interno». Puoi scoprire di poter facilmente immaginare che lei faccia *un piccolo* progresso in ciò – per esempio, immaginando dei trucchi tramite i quali sarebbe in grado di dire che un colore, qualunque esso sia, *non* è giallo, o *non* è rosso. Come? Notando qualche reazione specifica e rilevante che il suo cervello avrebbe solo per il giallo o solo per il rosso. Ma se le concedi anche un piccolo spiraglio per entrare nel suo spazio cromatico in questa maniera, dovrai concludere che possa far leva su questi dettagli per progredire verso una completa conoscenza, poiché non conosce solo le reazioni *rilevanti*, le conosce tutte.

Ricordati il cartoncino di Julius e Ethel Rosenberg, che era stato trasformato in un rilevatore di *M*. Ora immagina la loro

sorpresa se un impostore si presentasse con un pezzo che si accoppiasse bene senza essere l'originale. «Impossibile!» esclamerebbero. «Non impossibile», dice l'impostore, «solo difficile. Avevo *tutte le informazioni* necessarie per ricostruire un rilevatore di *M*, e per fare un'altra cosa con la proprietà formale *M*». Mary aveva sufficienti informazioni (nel caso originale, se immaginato correttamente) per capire esattamente cosa fossero i suoi rilevatori del rosso e i suoi rilevatori del blu, e quindi per identificarli in anticipo. Non è il modo usuale per arrivare a conoscere i colori, ma Mary non è una persona usuale.

Lo so che questo non soddisferà molti dei fans filosofici di Mary, e che ci sarebbe molto altro da dire, ma – e questo è il punto della mia argomentazione – l'effettiva dimostrazione deve svolgersi in un'arena che è ben distante dall'esempio di Jackson, che è un classico untore della Sindrome dei Filosofi: scambiare una carenza dell'immaginazione per l'intuizione di una necessità. Alcuni dei filosofi che hanno trattato il caso di Mary forse non si preoccupano di averlo immaginato erroneamente, poiché lo hanno usato semplicemente come trampolino per delle discussioni che gettano luce su vari argomenti di per sé interessanti e importanti. Non mi soffermerò qui su tali argomenti, poiché mi interessa esaminare direttamente la conclusione che Jackson stesso trae dal suo esempio: le esperienze visive hanno qualia che sono «epifenomenici».

Il termine «epifenomeno» è in uso comune oggi sia in filosofia sia in psicologia (e in altre scienze cognitive). È usato presupponendo che il suo significato sia familiare e comunemente accettato, quando in realtà i filosofi e gli scienziati cognitivi usano il termine con significati *completamente* differenti – questa stranezza mi sembra che sia resa ancora più strana dal fatto che sebbene io abbia più volte indicato questo problema, nessuno sembra preoccuparsene. Poiché «l'epifenomenalismo» sembra spesso l'ultimo porto sicuro per i qualia, e poiché l'apparenza di sicurezza è dovuta interamente alla confusione tra questi due significati, io devo diventare una megera e mettere sulla difensiva quelli che usano questo termine.

Secondo lo *Shorter Oxford English Dictionary*, il termine «epifenomeno» è apparso per la prima volta nel 1706 come un termine di patologia, «manifestazione o sintomo secondario». Il biologo evoluzionista Thomas Huxley (1874) è stato probabilmente lo scrittore che ha esteso il termine all'uso corrente in psicologia, dove significa proprietà *non-funzionale* o collaterale. Huxley ha usato il termine nella sua discussione dell'evoluzione

della coscienza e nella sua tesi che le proprietà epifenomeniche («come il fischio del motore a vapore») non possono essere spiegate dalla selezione naturale.

Ecco un esempio chiaro di questo uso della parola:

Perché le persone che stanno pensando intensamente si mordono le labbra e battono i piedi? Queste azioni sono semplici epifenomeni che accompagnano il nucleo centrale del processo del sentire e del pensare o potrebbero essere esse stesse parti di questi processi? (Zajonc e Markus, 1984, p. 74.)

Si noti che gli autori intendono asserire che queste azioni, sebbene siano perfettamente rilevabili, non giocano nessun ruolo programmato, nessun ruolo che potenzia le prestazioni, nei processi del sentire e del pensare; sono non funzionali. Nello stesso spirito, il ronzio del calcolatore è epifenomenico, così come la tua ombra quando ti fai un tè. Gli epifenomeni sono solo prodotti collaterali, ma in quanto tali sono prodotti con parecchi effetti sul mondo: battere i piedi produce un rumore registrabile, e la tua ombra ha degli effetti su una lastra fotografica, per non parlare del leggero raffreddamento delle superfici su cui si proietta.

L'usuale significato filosofico è differente: «*x* è epifenomenico» significa «*x* è un effetto ma esso stesso non ha alcun effetto sul mondo fisico». (Vedi Broad, 1925, p. 118, per la definizione che inaugura, o in ogni caso fissa, l'uso filosofico.) Questi significati sono davvero così differenti? Sì, così differenti come i significati di *assassinio* e *morte*. Il significato filosofico è più forte: qualunque cosa che non abbia alcun effetto nel mondo fisico sicuramente non ha effetti sulla funzione di alcunché, ma l'inverso non è vero, come l'esempio di Zajonc e Markus rende ovvio.

A ben vedere, il significato filosofico è troppo forte; comporta un concetto di nessuna utilità (Harman, 1990; Fox, 1989). Poiché *x* non ha effetti fisici (secondo la sua definizione), nessuno strumento può rilevare la presenza di *x*, sia direttamente che indirettamente; gli avvenimenti del mondo non sono minimamente influenzati dalla presenza o dall'assenza di *x*. Come potrebbe mai esserci, allora, una qualche ragione empirica per affermare la presenza di *x*? Supponiamo, per esempio, che Otto insista nell'affermare che lui (ad esempio) ha qualia epifenomenici. Perché dice questo? Non perché essi abbiano qualche effetto su di lui, lo guidino o lo avvertano mentre fa la

sua affermazione. Per la stessa definizione di epifenomeno (nel senso filosofico), le più sincere affermazioni di Otto sull'esistenza dei suoi epifenomeni *non possono* essere una prova per lui né per nessun altro che essi esistano davvero, poiché egli direbbe esattamente la stessa cosa anche se non ci fossero. Ma forse Otto dispone di qualche prova «interna»?

Qui c'è una scappatoia, ma non molto attraente. Gli epifenomeni, ci si ricordi, sono definiti come privi di effetti sul mondo *fisico*. Se Otto vuole abbracciare completamente il dualismo, egli può sostenere che i suoi qualia epifenomenici non hanno effetti sul mondo fisico, ma li hanno nel suo mondo mentale (non fisico). (Broad, 1925, ha chiuso questa scappatoia per definizione, ma è sempre a disposizione per chi la desidera.) Per esempio, *essi causano alcune delle sue credenze (non fisiche)*, come la credenza di avere qualia epifenomenici. Ma questa è solo una fuga temporanea dalla difficoltà. Perché ora, se non si vuole contraddire, le sue credenze non possono, a loro volta, avere alcun effetto sul mondo fisico. Se perdesse improvvisamente i suoi qualia epifenomenici, non crederebbe più di averli, ma continuerebbe imperterrito *a dire* di averli. Solo che non crederebbe a quello che dice! (Né potrebbe dirci che non crede a quello che dice o fare qualunque cosa che potesse segnalarci che non crede più a quello che stava dicendo.) Così l'unico modo in cui Otto potrebbe «giustificare» la sua credenza negli epifenomeni sarebbe ritirandosi in un mondo solipsistico in cui ci sia solo lui, le sue credenze e i suoi qualia, tagliato fuori da ogni effetto sul mondo. Lungi dall'essere un modo «sicuro» di essere un materialista e di avere anche i propri qualia, questo è nella migliore delle ipotesi un modo per sostenere il solipsismo più radicale, tagliando fuori la propria mente – le proprie credenze e le proprie esperienze – da ogni commercio con il mondo materiale.

Se i qualia sono epifenomenici nell'usuale senso filosofico, la loro presenza non può spiegare il modo in cui le cose si verificano (nel mondo materiale) poiché, per definizione, le cose accadrebbero esattamente nello stesso modo senza di loro. Non ci potrebbe essere, allora, una ragione empirica per credere negli epifenomeni. Potrebbe esserci un altro tipo di ragione per asserire la loro esistenza? Che tipo di ragione? Una ragione *a priori*, presumibilmente. Ma quale? Nessuno ne ha mai presentata una – buona, cattiva o indifferente – che io sappia. Se qualcuno vuole obiettare che sto comportandomi da «verificazionista» rispetto a questi epifenomeni, rispondo: non siamo

tutti verificazionisti rispetto a *questo* tipo di asserzioni? Si consideri, per esempio, l'ipotesi che ci siano quattordici folletti epifenomenici in ogni cilindro di un motore a combustione interna. Questi folletti non hanno massa, né energia, né proprietà fisiche; non fanno andare il motore più regolarmente né più irregolarmente, né più veloce né più lentamente. Non c'è *e non ci può essere* nessuna prova empirica della loro presenza, e nessun modo empirico in linea di principio per distinguere questa ipotesi dalle sue rivali: ci sono tredici o quindici... folletti. Tramite quale principio difendiamo il nostro totale rifiuto di una tale insensatezza? Un principio verificazionista o solo semplice buon senso?

Ah, ma c'è una differenza! [dice Otto]. Non c'è nessun motivo reale per prendere sul serio l'ipotesi di questi folletti. Li hai inventati solo lì per lì. I qualia, invece, hanno una lunga esistenza e giocano un ruolo fondamentale nei nostri schemi concettuali!

E se un popolo di ignoranti avesse pensato per varie generazioni che siano i folletti a far andare le loro automobili e ora, spinti dal progresso della scienza, si rifugiassero nell'affermazione che i folletti esistono sul serio, ma sono epifenomenici? Sbaglieremmo se liquidassimo le loro «ipotesi» seduta stante? Quale che sia il principio su cui ci basiamo quando buttiamo dalla finestra siffatte insensatezze, esso è sufficiente a liquidare la dottrina che i qualia siano epifenomenici in questo senso filosofico. Queste non sono posizioni che meritano di essere discusse seriamente.

È difficile credere che i filosofi che hanno recentemente descritto la loro posizione come epifenomenalismo possano fare un errore tanto desolante. Forse stanno asserendo che i qualia sono epifenomenici nel senso di Huxley? I qualia, secondo questa interpretazione, *sono* effetti fisici e *hanno* effetti fisici; solamente non sono funzionali. Ogni materialista dovrebbe essere ben contento di ammettere come vera una tale ipotesi – se identifichiamo i qualia con le disposizioni a reagire, per esempio. Come abbiamo notato nella discussione sul godimento, anche se alcune propensioni o predisposizioni nel nostro spazio qualitativo sono funzionali – o erano funzionali – altre sono solo semplici fatti casuali. Perché non mi piacciono i broccoli? Probabilmente non c'è alcuna ragione; la mia disposizione reattiva negativa è puramente epifenomenica, un pro-

dotto collaterale del mio cablaggio senza alcuna significanza. Non ha una funzione, ma ha molti effetti. In ogni sistema progettato, alcune proprietà sono cruciali, mentre altre sono più o meno revisionabili *ad libitum*. Ogni cosa deve essere in un modo o nell'altro, ma spesso il modo non conta. La leva del cambio di un'automobile forse deve essere di una certa lunghezza e di una certa solidità, ma se la sua sezione è rotonda o quadrata o ovale è una proprietà epifenomenica, nel senso di Huxley. Nei sistemi CADBLIND che abbiamo immaginato nel capitolo 10, lo schema particolare di codificazione tramite numeri era epifenomenico. Potremmo «invertirlo» (usando numeri negativi o moltiplicando tutti i valori per qualche costante) senza produrre alcuna differenza *funzionale* nei suoi poteri di elaborazione dell'informazione. Una tale inversione potrebbe essere non rilevabile da un'ispezione superficiale, e potrebbe essere irrilevabile *dal sistema*, ma non sarebbe epifenomenica nel senso filosofico. Ci sarebbe una miriade di differenze di piccolo voltaggio nei registri della memoria che contengono i numeri differenti, per esempio.

Se pensiamo a tutte le proprietà del nostro sistema nervoso che ci permettono di vedere, ascoltare, odorare, gustare e toccare le cose, possiamo dividerle, approssimativamente, in quelle che giocano un ruolo davvero cruciale nel mediare l'elaborazione dell'informazione e quelle epifenomeniche che sono più o meno revisionabili *ad libitum*, come il sistema di codificazione del colore nel CADBLIND. Quando un filosofo suppone che i qualia siano proprietà epifenomeniche degli stati cerebrali, ciò potrebbe significare che i qualia si rivelino essere delle variazioni locali nel calore generato dal metabolismo neuronale. Questo non può essere ciò che gli epifenomenalisti hanno in mente, no? Se è così, allora i qualia come epifenomeni non rappresentano una sfida per il materialismo.

È giunto il momento di passare l'onere della prova direttamente a quelli che insistono nell'usare il termine. Il senso filosofico del termine è semplicemente ridicolo; il senso di Huxley è relativamente chiaro e non problematico – e non rilevante per gli argomenti filosofici. Nessun altro senso del termine è in circolazione. Così se qualcuno pretende di sostenere una varietà di epifenomenalismo, prova a essere educato ma domanda: «Di che diamine stai parlando?».

Si noti, tra parentesi, che questo equivoco tra i due sensi di «epifenomenico» influenza anche la discussione sugli zombi. Lo zombi del filosofo, si ricorderà, è comportamentalmente indi-

stinguibile da un normale essere umano, ma non è cosciente. Non si prova nulla ad essere uno zombi; che ad esserlo si provi qualcosa sembra soltanto agli osservatori (incluso se stesso, come abbiamo visto nel capitolo precedente). Ora, a ciò si può dare un'interpretazione debole e una forte, a seconda di come trattiamo questa indistinguibilità da parte degli osservatori. Se dovessimo dichiarare che, *in linea di principio*, uno zombi è indistinguibile da una persona cosciente, allora ci troveremmo a dire che la coscienza autentica è epifenomenica *nel senso ridicolo*. Questo è semplicemente sciocco. Così potremmo dire invece che la coscienza forse è epifenomenica nel senso di Huxley: benché ci sia qualche modo per distinguere gli zombi dalle persone normali (chissà, forse gli zombi hanno i cervelli verdi), la differenza non si rivela come una differenza funzionale *agli occhi degli osservatori*. In altre parole, i corpi umani con cervelli verdi non albergano osservatori, a differenza degli altri. In base a questa ipotesi saremmo in grado di distinguere in linea di principio i corpi abitati da quelli non abitati controllando il colore dei cervelli. Anche questo è sciocco, naturalmente, e pericolosamente sciocco, perché riecheggia la sorta di pregiudizi completamente immotivati che hanno negato la piena personalità alla gente sulla base del colore della loro pelle. È tempo di riconoscere l'idea della possibilità degli zombi per quello che è: non un'idea filosofica seria, ma un'assurda e ignobile reliquia di antichi pregiudizi. Forse le donne non sono realmente coscienti! Forse gli ebrei! Che pericolosa insensatezza. Come dice Shylock, attirando la nostra attenzione, piuttosto appropriatamente, sui criteri «meramente comportamentali»:

> Un ebreo non ha forse occhi? Un ebreo non ha forse mani, organi, dimensioni, sensi, affetti, passioni; non è nutrito dallo stesso cibo, ferito dalle stesse armi, soggetto alle stesse malattie, curato dagli stessi rimedi, riscaldato e agghiacciato dagli stessi inverni e dalle stesse estati, come un cristiano? Se ci pungete, non sanguiniamo? Se ci fate il solletico, non ridiamo? Se ci avvelenate, non moriamo?

C'è un altro modo di rivolgersi alla possibilità degli zombi, e per certi aspetti credo che sia più soddisfacente. Gli zombi sono possibili? Non solo sono possibili, sono reali. Siamo tutti

zombi.[6] Nessuno è cosciente – non nel modo sistematicamente misterioso che sorregge dottrine come l'epifenomenalismo! Non posso dimostrare che nessuna specie di coscienza del genere esista. Non posso neanche dimostrare che i folletti non esistano. La cosa migliore che posso fare è mostrare che non c'è alcun motivo rispettabile per credere in essa.

6. *Di nuovo sulla mia sedia a dondolo*

Nel secondo paragrafo del secondo capitolo, ho delineato il compito di spiegare la coscienza ricordando un episodio della mia esperienza cosciente mentre dondolandomi sulla mia sedia osservavo fuori dalla finestra un magnifico giorno di primavera. Ritorniamo a quel passo e vediamo come viene trattato dalla teoria che ho sviluppato. Ecco il testo:

> In quel giorno di una precoce primavera, un fascio verde-dorato di luce solare penetrava dalla finestra e attraverso un velo di germogli verdi era ancora possibile vedere le migliaia di rami e ramoscelli dell'acero del mio giardino: il tutto formava un intricato ed elegante merletto. Il vetro della finestra è antico ed è attraversato da una corrugazione appena rilevabile, mentre mi dondolavo avanti e indietro questa imperfezione del vetro provocava un'onda di saltelli sincronizzati che si spostava avanti e indietro nel fascio dei rami, era un movimento regolare sovrapposto con notevole chiarezza allo scintillio più caotico dei rami e ramoscelli nel vento.
>
> Poi notai che questo metronomo visivo nei rami dell'albero era legato al ritmo del concerto grosso di Vivaldi che stavo ascoltando come «musica di sottofondo» per la mia lettura [...]. Il mio pensiero cosciente, e specialmente la gioia che provavo per quella combinazione di luce solare, di solari violini di Vivaldi, di rami ondulanti – più il piacere delle mie riflessioni su ciò – come poteva *tutto questo* essere solo qualcosa di fisico che avveniva nel mio cervello? Come poteva qualsiasi combinazione di avvenimenti elettro-chimici nel mio cervello equivalere in qualche modo alla piacevolezza di quelle centinaia di ramoscelli che si inginocchiavano di fronte a me a tempo di musica? Come poteva qualche evento che elaborava informazioni nel mio cervello *essere* il calore delicato della luce solare che mi accarezzava?... Sembra davvero impossibile.

Poiché ho incoraggiato tutti noi ad essere eterofenomenologi, non posso certo escludermi e devo essere contento tanto di essere il praticante quanto il soggetto, quindi applichiamo la mia teoria a me stesso. Come eterofenomenologi, il nostro

compito è di prendere questo testo, interpretarlo e poi correlare gli oggetti che risultano dal mondo eterofenomenologico di Dennett con gli eventi che accadono in quel momento nel cervello di Dennett.

Poiché il testo è stato prodotto qualche settimana o qualche mese dopo gli eventi di cui parla, possiamo essere sicuri che sia stato abbreviato, non solo dalle compressioni editoriali deliberate dell'autore, ma anche dall'inesorabile processo di riduzione della memoria nel tempo. Se l'avessimo sondato prima – se l'autore avesse preso un registratore in quel momento e avesse prodotto il testo seduta stante – sarebbe stato sicuramente differente. Non solo più ricco nei dettagli, e più confuso, ma anche, naturalmente, riplasmato e ridiretto dalla reazione dell'autore al processo stesso di creazione del testo – all'ascolto del suono delle sue parole invece della riflessione silenziosa. Parlare a voce alta, come ogni conferenziere sa, rivela spesso implicazioni (e particolarmente problemi) nel proprio messaggio che vengono eluse quando ci si impegna in un soliloquio silenzioso.

Così com'è, il testo ritrae solo una porzione (e senza dubbio una porzione idealizzata) del contenuto della coscienza dell'autore. Dobbiamo stare attenti, comunque, a non supporre che le «parti lasciate fuori» dal testo dato siano tutte «realmente presenti» in ciò che potrebbe essere chiamato il flusso della coscienza dell'autore. Non dobbiamo fare l'errore di supporre che ci siano fatti – non recuperabili ma reali – che ci possano aiutare a stabilire quali contenuti precisi fossero coscienti o no in quel momento. E in particolare, non dobbiamo supporre che quando guardava fuori della finestra, egli «abbia inghiottito tutto» in un fantastico boccone mentale – anche se questo è ciò che il suo testo ritrae. A lui è sembrato, secondo il testo, come se la sua mente – il suo campo visivo – fosse invasa da intricati dettagli di germogli verdi-dorati e di rami saltellanti, ma benché questo è come sembrava, è un'illusione. Nessun «pieno» del genere è mai giunto alla sua mente; il pieno è rimasto fuori nel mondo dove non aveva bisogno di essere *rappresentato*, ma poteva solo essere. Quando noi ci meravigliamo, in quei momenti di autocoscienza potenziata, della sorprendente ricchezza della nostra esperienza cosciente, la ricchezza di cui ci meravigliamo è in realtà la ricchezza del mondo esterno, in tutti i *suoi* affascinanti dettagli. Non «entra» nelle nostre menti coscienti, ma è semplicemente disponibile.

E che dire di tutti i rami e i germogli che si increspano all'unisono? I rami fuori sull'albero non si increspano di sicuro,

poiché l'increspatura è dovuta alla corrugazione del vetro, ma questo non significa che tutta l'increspatura doveva avvenire nella mente o nel cervello dell'autore, solo che avveniva al di qua del vetro che l'ha causata. Se qualcuno avesse filmato le immagini cangianti sulla retina dell'autore, avrebbe trovato lì l'increspatura, proprio come in un film, ma questo è senza dubbio anche il posto dove tutte le increspature si sono fermate; quello che accadde dopo le retine fu solo il suo riconoscimento che c'era, come dice il testo, un'onda meravigliosa di increspature sincronizzate di cui poteva avere esperienza. Egli vide le increspature, e vide la loro estensione, nello stesso modo in cui vedresti tutte le Marilyne sulla carta da parati. E poiché le sue retine erano provviste di una persistente dose di increspature, se gli fosse andato di esaminarle ulteriormente, ci *sarebbero stati* più dettagli nelle Versioni Multiple di cui il nostro testo è tutto ciò che rimane.

C'erano molti altri dettagli sui quali l'autore si sarebbe potuto focalizzare, ma non lo ha fatto. C'è una miriade di verità autentiche ma non recuperabili su quali di questi dettagli sia stato discriminato, dove e quando, dai vari sistemi del suo cervello, ma la somma totale di queste verità non risolve interrogativi come: «Di quali di questi dettagli egli è stato realmente e sicuramente cosciente (ma si è dimenticato prima di aver prodotto il suo testo)?» o «Quali erano realmente e sicuramente sullo "sfondo" della sua coscienza (sebbene non li abbia focalizzati in quel momento)?». La nostra tendenza a supporre che ci *debba* essere una verità che risolva tali questioni è come la supposizione di un lettore ingenuo che pensi che ci *debba* essere una risposta a domande come: «Sherlock Holmes ha mangiato uova a colazione il giorno che il dottor Watson l'ha incontrato?». Conan Doyle *potrebbe aver* messo quel dettaglio nel testo, ma non lo ha fatto, e dal momento che non lo ha fatto non c'è una verità sull'appartenenza o meno di quelle uova al *mondo fittizio di Sherlock Holmes*. Anche se Conan Doyle *ha pensato* che Holmes aveva mangiato uova quella mattina, anche se in una versione precedente Holmes *viene rappresentato con parole esplicite* mentre mangia uova quella mattina, non ci sono affatto delle verità sul mondo fittizio di Sherlock Holmes, il mondo costituito dal testo pubblicato che noi effettivamente abbiamo, che ci possano chiarire se egli abbia o no mangiato uova a colazione.

Il testo che abbiamo ricevuto da Dennett non era «scritto nel suo cervello» nell'intervallo di tempo tra l'episodio della sedia a dondolo e il momento in cui è stato dattiloscritto in un

file di un programma di videoscrittura. La cura con cui egli ha seguito lo spettacolo che gli si presentava mentre si dondolava e la concomitante ripetizione di quei particolari che avevano attirato la sua attenzione hanno avuto l'effetto di fissare i contenuti di quei particolari in modo relativamente sicuro «nella memoria», ma questo effetto non deve essere visto come un immagazzinamento di una fotografia (o di una frase) o qualsiasi altra rappresentazione significativa del genere. Deve essere pensato, piuttosto, come qualcosa che rende più verosimile la ripresentazione parzialmente simile dell'attività, e quell'evento simile è quello che è avvenuto, possiamo presupporre, nell'occasione della scrittura, quando le parole/folletti nel suo cervello hanno formato delle coalizioni che hanno prodotto, per la prima volta, una successione di frasi. Ora, alcune delle cose che accaddero precedentemente, sulla sedia a dondolo, hanno senza dubbio arruolato effettive parole e frasi in inglese, e questa collaborazione precedente tra contenuti privi di parole e parole ha senza dubbio facilitato il recupero di alcune di quelle espressioni inglesi che si sono presentate al momento della scrittura.

Torniamo al mondo eterofenomenologico di quel testo. Che dire della gioia di cui parla? «...la combinazione di luce solare, di solari violini di Vivaldi, di rami ondulanti – più il piacere delle mie riflessioni su ciò...». Questo non può essere spiegato chiamando in causa i qualia intrinsecamente piacevoli della vista, dell'udito e del puro pensiero. L'idea che ci siano qualia del genere non fa altro che distrarci da ogni possibile approccio esplicativo, catturando la nostra attenzione come quella di un bebè quando gli si agita un dito di fronte agli occhi, inducendoci a fissare instupiditi «l'oggetto intrinseco» invece di darci da fare per costruire una descrizione dei meccanismi sottostanti e una spiegazione (che sia in un'ultima analisi evoluzionistica) del perché i meccanismi fanno quello che fanno.

Il godimento dell'autore è facilmente spiegabile dal fatto che tutta l'esperienza visiva è composta di attività dei circuiti neurali la cui attività stessa è per natura piacevole, non solo perché ci piace essere informati ma perché ci piacciono le maniere specifiche in cui veniamo informati. Il fatto che i germogli primaverili chiazzati di luce solare siano qualcosa che piace agli esseri umani non è sorprendente. Il fatto che a qualche essere umano piace anche guardare diapositive di batteri al microscopio, e ad altri fotografie di incidenti aerei è più strano, ma le sublimazioni e le perversioni del desiderio

crescono dalla stessa origine animale nel cablaggio dei nostri sistemi nervosi.

L'autore prosegue chiedendosi come sia possibile che «tutto ciò sia solo una combinazione di avvenimenti elettrochimici nel mio cervello». Come il suo stupore mette in luce, non sembra affatto così. O in ogni caso c'è stato un momento in cui egli ha pensato che non sembrava affatto essere solo una combinazione di eventi elettrochimici nel suo cervello. Ma i nostri successivi capitoli suggeriscono una replica: ma come pensi che sembrerebbe se *fosse* solo una combinazione di avvenimenti elettrochimici nel tuo cervello?[7] Non ci siamo costruiti delle basi sufficienti per concludere che con un cervello organizzato nella maniera in cui è organizzato il nostro cervello, questa è esattamente la sorta di mondo eterofenomenologico che ci aspetteremmo? Perché tali combinazioni di eventi elettrochimici nel cervello non dovrebbero avere proprio gli effetti che ci siamo proposti di spiegare?

(Parla l'autore): C'è ancora un enigma, comunque. Come vengo *io* a sapere tutto ciò? Come è possibile che *io* posso dirti tutto quello che avviene nella mia testa? La risposta all'enigma è semplice: *perché quello è ciò che io sono*. Perché un conoscitore e un descrittore di tali cose in tali termini è ciò che io sono. La *mia* esistenza è spiegata dal fatto che ci sono queste capacità in questo corpo.

Questa idea, l'idea di un Sé come di un Centro di Gravità Narrativa, è quella che siamo infine pronti ad esaminare. È certamente un'idea matura per i tempi. Si immaginino le mie emozioni contrastanti quando ho scoperto che prima che potessi vedere la mia versione di essa adeguatamente pubblicata in un libro,[8] essa era già stata presa in giro in un romanzo, *Ottimo lavoro, professore!* di David Lodge (1988). È evidentemente un tema caldo tra i decostruzionisti:

Secondo Robyn (o, più esattamente, secondo quegli scrittori che hanno influenzato il suo pensiero su questi argomenti), non esiste niente come il «Sé» su cui sono fondati il capitalismo e il romanzo classico – cioè, quell'unica e determinata essenza o anima che costituisce l'identità di una persona; c'è solo la posizione del soggetto attorno alla quale viene tessuta una ragnatela infinita di discorsi: i discorsi di potere, sesso, famiglia, scienza, religione, della poesia, ecc. E, alla stessa stregua, non esiste l'autore, cioè colui che inventa un lavoro di fantasia dal nulla... nelle parole famose di Derrida... «*il n'y a pas de hors-*

texte», «non c'è niente al di fuori del testo». Non ci sono origini, ma solo una produzione, e noi produciamo i nostri «sé» nel linguaggio. Non «*sei quello che mangi*» ma «*sei quello che dici*» o, piuttosto, «*sei quello che ti dice*» è la base assiomatica della filosofia di Robyn, ciò che lei, se le si chiedesse di dargli un nome, chiamerebbe «materialismo semiotico».

Materialismo semiotico? Dovrei chiamarlo così? A parte le allusioni al capitalismo e al romanzo classico, su cui non mi pronuncio, questo passaggio satirico è un'eccellente parodia della posizione che sto per presentare. (Come tutte le parodie, esagera; non direi che non c'è *nulla* al di fuori del testo. Ci sono, per esempio, tutte le biblioteche, gli edifici, i corpi, i batteri...)

Robyn ed io pensiamo in modo simile – e naturalmente siamo entrambi, secondo le nostre spiegazioni, una specie di personaggio fittizio, sebbene di una specie leggermente differente.

13

LA REALTÀ DEI SÉ

Se immaginiamo una macchina costruita in modo che pensi, senta, percepisca, si potrà concepire che venga ingrandita conservando le medesime proporzioni, in modo che vi si possa entrare come in un mulino. Ciò fatto, nel visitarla internamente non si troverà altro che pezzi, i quali si spingono scambievolmente e mai alcuna cosa che possa spiegare una percezione.

GOTTFRIED WILHELM LEIBNIZ (1646-1716), *Monadologie*

Per parte mia, quando mi addentro più profondamente in ciò che chiamo *me stesso*, mi imbatto sempre in una particolare percezione: di caldo o di freddo, di luce o di oscurità, di amore o di odio, di dolore o di piacere. Non riesco mai a sorprendere *me stesso* senza una percezione e a cogliervi altro che la percezione. [...] Se qualcuno, dopo una seria e spregiudicata riflessione, crede di avere una nozione differente di *se stesso*, dichiaro che non posso seguitar a ragionare con lui. Tutt'al più, gli potrei concedere che può aver ragione come l'ho io, e che in questo punto siamo essenzialmente differenti: egli forse percepisce qualcosa di semplice e di continuo, che chiama *se stesso*, mentre io son certo che in me un tal principio non esiste.

DAVID HUME (1739)

Fin dagli albori della scienza moderna nel diciassettesimo secolo, si è ritenuto quasi unanimemente che il sé, qualunque cosa esso sia, debba essere invisibile sia allo scrutinio degli strumenti scientifici che all'introspezione. Ad alcuni, ciò ha suggerito che il sé sia un'anima non fisica, uno spettro nella macchina. Ad altri, ha suggerito che il sé non esista affatto, che sia solo un figmento creato da immaginazioni metafisicamente infervorate. Ad altri ancora, ha suggerito soltanto che il sé sia in un certo senso una sorta di astrazione, qualcosa la cui esistenza non è minimamente contraddetta dalla sua invisibilità. Dopo tutto, si potrebbe dire, un centro di gravità è altrettanto invisibile – e altrettanto reale. È però sufficientemente reale?

Si può far in modo che la questione sulla reale esistenza dei

sé appaia ridicolmente facile da risolvere, in entrambe le direzioni. *Noi* esistiamo? Naturalmente! La domanda presuppone la sua risposta. (Dopo tutto, chi è questo *io* che ha cercato invano un sé, secondo Hume?) Ci sono delle entità, o *nei* nostri cervelli o *oltre e al di là* di essi, che controllano i nostri corpi, pensano i nostri pensieri, prendono le nostre decisioni? Naturalmente no! Una idea siffatta è o un'idiozia empirica (il «neurone pontificio» di James) o uno sproloquio metafisico (lo «spettro nella macchina» di Ryle). Quando una domanda semplice ottiene due risposte, «Ovviamente sì!» e «Ovviamente no!», vale la pena considerare una posizione a metà strada (Dennett, 1991a), anche se è destinata ad apparire non intuitiva ad antrambe le parti – ognuno afferma che nega uno o l'altro dei fatti ovvi!

1. *Come gli esseri umani tessono* un sé*

> Inoltre sembrava che impiegassero la maggior parte del tempo a mangiare e bere e andare alle feste, e Frensic, il cui aspetto tendeva a limitare i suoi piaceri sensuali al mettere le cose in se stesso piuttosto che in altre persone, era una specie di buongustaio.
>
> Tom Sharpe (1977)

Il romanziere Tom Sharpe suggerisce, in questo brano divertente ma sconvolgente, che, a ben vedere, tutti i piaceri sensuali consistono nel giocare con i propri confini, o quelli di qualcun altro; egli ha colto qualcosa d'importante – se non tutta la verità, almeno parte di essa.

La gente ha un sé. E i cani? E le aragoste? Se i sé sono davvero qualcosa, allora esistono. *Adesso* ci sono i sé. C'è stato un tempo, migliaia (o milioni o miliardi) di anni fa, quando non ce n'era nessuno – almeno su questo pianeta. Allora è logicamente possibile immaginare una storia che dica *come si è giunti all'esistenza* di creature con un sé. Questa storia dovrà logicamente raccontare un processo (o una serie di processi) che implica le attività o i comportamenti di entità che non *hanno* ancora un sé – o non *sono* ancora un sé – ma che conduce alla fine, come un nuovo prodotto, a esseri che sono, o hanno, un sé.

* Il verbo inglese *to spin*, estremamente importante per la sua pregnanza in questo capitolo, è stato tradotto con l'italiano *tessere*, si tenga presente però che significa anche *filare, secernere, raccontare*. (N.d.T.)

Nel capitolo 7, abbiamo visto come la nascita delle ragioni fosse anche la nascita dei confini, il confine tra «me» e «il resto del mondo», una distinzione che anche la più umile ameba deve fare, nel suo modo cieco e inconsapevole. Questa inclinazione minima a distinguere sé dagli altri per proteggere *se stessi* costituisce il sé biologico, e perfino un sé così semplice non è un'entità concreta ma solo un'astrazione, un principio organizzativo. Inoltre i confini del sé biologico sono porosi e indefiniti – un altro esempio di come Madre Natura tollera un «errore» se non costa troppo.

All'interno delle pareti del corpo umano ci sono molti, molti intrusi: dai batteri e virus ai microscopici acari che vivono come rocciatori nelle nicchie ecologiche della nostra pelle e del nostro cuoio capelluto, fino ai parassiti più grandi – le orribili tenie, per esempio. Tutti questi piccoli intrusi proteggono i loro propri diritti, ma alcuni di loro, come i batteri che popolano il nostro sistema digestivo e senza i quali moriremmo, sono membri di una squadra altrettanto indispensabile ai nostri sforzi di autoconservazione di quella formata dagli anticorpi del nostro sistema immunitario. (Se la teoria della biologa Lynn Margulis [1970] è corretta, i mitocondri che svolgono i loro compiti in quasi tutte le cellule del nostro corpo sono i discendenti dei batteri con i quali «noi» abbiamo unito le forze due miliardi di anni fa.) Altri intrusi sono dei parassiti tollerati – evidentemente, non vale la pena di sfrattarli – altri ancora sono effettivamente dei nemici interni, mortali se non vengono sradicati.

Questo principio biologico fondamentale di distinzione tra sé e il mondo, interno e esterno, produce qualche eco osservabile nelle sfere più alte della nostra psicologia. Gli psicologi Paul Rozin e April Fallon (1987) hanno mostrato in una serie affascinante di esperimenti sulla natura del *disgusto* che c'è una tendenza nascosta, potente e non riconosciuta, a resistere ciecamente a certi atti che, considerati razionalmente, non dovrebbero infastidirci. Per esempio, ingoia per piacere la saliva che hai ora nella bocca. Questo atto non ti riempie di repulsione. Ma supponiamo che ti avessi chiesto di prendere un bicchiere pulito, di sputarci dentro e *poi* inghiottire la saliva dal bicchiere. Disgustoso! Ma perché? Sembra avere a che fare con la nostra impressione che quando qualcosa è fuori del nostro corpo non fa più parte di noi – diventa estranea e sospetta – ha rinunciato alla sua cittadinanza e diventa qualcosa da respingere.

Gli attraversamenti di confine sono quindi o momenti di ansietà o, come indicato da Sharpe, qualcosa di cui si gode in

modo particolare. Molte specie hanno sviluppato costruzioni notevoli per estendere i loro confini territoriali, o per rendere più difficili gli attraversamenti dannosi o per facilitare quelli positivi. I castori costruiscono dighe e i ragni filano le ragnatele, per esempio. Quando un ragno tesse la sua tela, non deve comprendere cosa sta facendo; Madre Natura ha fornito al suo minuscolo cervello le procedure necessarie per portare a termine questo compito ingegneristico biologicamente essenziale. Esperimenti con i castori mostrano che perfino le loro pratiche ingegneristiche meravigliosamente efficienti sono in larga parte un prodotto di pulsioni e inclinazioni innate che essi non devono capire per trarne un beneficio. I castori apprendono, e possono anche insegnare ad altri castori, ma sono fondamentalmente guidati da potenti meccanismi innati che controllano ciò che il comportamentista B.F. Skinner chiamava il rinforzo negativo. Un castoro ricercherà sempre, piuttosto affannosamente, qualcosa – qualunque cosa – per fermare il suono dell'acqua che scorre; in un esperimento un castoro non ha trovato nulla di meglio per darsi sollievo che ricoprire di fango l'altoparlante da cui proveniva il gorgoglio registrato! (Wilsson, 1974.)

Il castoro protegge il suo confine esterno con ramoscelli e fango e uno dei suoi confini interni con la pelliccia. La lumaca raccoglie il calcio dal suo cibo e lo usa per secernere una conchiglia; il paguro si procura una conchiglia di calcio già fatta, prendendo quella scartata da un'altra creatura, evitando schizzinosamente il processo di ingestione e di secrezione. La differenza non è fondamentale, secondo Richard Dawkins, il quale mette in luce che, in entrambi i casi, il risultato, che egli chiama *il fenotipo esteso* (1982), è parte dell'equipaggiamento biologico fondamentale degli individui che sono sottomessi alle forze selettive che guidano l'evoluzione.

La definizione di un fenotipo esteso non solo si estende oltre il confine «naturale» degli individui fino a comprendere gli equipaggiamenti esterni come le conchiglie (e gli equipaggiamenti interni come i batteri residenti), ma spesso include altri individui della stessa specie. I castori non possono farcela da soli, è necessario un lavoro di squadra per costruire anche una sola diga. Le termiti devono associarsi a milioni per costruire i loro castelli.

E si considerino le stupefacenti costruzioni architettoniche dell'uccello giardiniere australiano (Borgia, 1986). I maschi costruiscono elaborate «capanne», templi per il corteggiamento

con grandi navate centrali, riccamente decorati con oggetti brillanti e colorati – soprattutto blu scuro, tra cui tappi di bottiglia, pezzi di vetro colorato e altri manufatti umani – che sono raccolti anche da molto lontano e attentamente assemblati nelle capanne nel modo migliore per impressionare la femmina corteggiata. L'uccello giardiniere, come il ragno, non deve realmente capire cosa sta facendo; si scopre semplicemente intento al lavoro, non sapendo perché, creando un edificio che è cruciale al suo successo come uccello giardiniere.

Ma le costruzioni più strane e più meravigliose dell'intero mondo animale sono le stupefacenti e intricate costruzioni realizzate dal primate *homo sapiens*. Ogni individuo normale di questa specie crea un *sé*. Con il suo cervello tesse una rete di parole e di atti, e, come altre creature, non deve sapere cosa sta facendo; semplicemente, lo fa. Questa rete lo protegge, come la conchiglia della lumaca; e gli fornisce i mezzi di sostentamento, come la rete del ragno; e incrementa le sue fortune sessuali, come la capanna dell'uccello giardiniere. A differenza del ragno, un individuo umano non *secerne* semplicemente la sua rete; più simile a un castoro, lavora sodo per procurarsi il materiale con cui costruire la sua fortezza protettiva. Come l'uccello giardiniere, si appropria di molti oggetti trovati che piacciono a lui – o alla sua compagna – compresi quelli che sono stati progettati da altri per altri scopi.

Questa «ragnatela di discorsi», come viene chiamata da Robyn, alla fine del capitolo precedente, è un prodotto biologico come qualunque altra costruzione trovata nel mondo animale. Spogliato di essa, un essere umano è altrettanto incompleto di un uccello senza piume, di una tartaruga senza conchiglia. (Anche i vestiti sono parte del fenotipo esteso dell'*homo sapiens* in quasi tutte le nicchie abitate da questa specie. Un'enciclopedia illustrata di zoologia non avrebbe una giustificazione maggiore a raffigurare l'*homo sapiens* nudo di quanto non ne abbia a raffigurare l'*ursus arctos* – l'orso bruno – con un vestito da pagliaccio e a cavalcioni di una bicicletta.)

L'organizzazione di una colonia di termiti è così meravigliosa che alcuni osservatori hanno pensato che ogni colonia deve avere un'anima (Marais, 1937). Noi ora comprendiamo che la sua organizzazione è semplicemente il risultato di un milione di piccoli agenti semi-indipendenti, ognuno un automa che svolge il suo lavoro. L'organizzazione di un sé umano è così meravigliosa che molti osservatori hanno pensato che anche

ogni essere umano abbia un'anima: un benevolo Dittatore che governa dal Quartier Generale.

In ogni alveare e in ogni termitaio c'è, sicuramente, un'ape regina e una termite regina, ma questi individui sono più passivi che attivi, assomigliano di più ai gioielli della corona da proteggere che ai capi delle forze di protezione – infatti il loro appellativo regale è più adatto oggi che nel passato, giacché assomigliano di più alla regina Elisabetta II che alla regina Elisabetta I. Non c'è un'ape Margaret Thatcher, una termite George Bush, un ufficio presidenziale nel formicaio.

I nostri sé, i nostri complessi *sé egoici*, mostrano la stessa permeabilità e flessibilità di confini dei *sé* più semplici delle altre creature? Espandiamo i nostri confini personali – i confini dei nostri sé – per comprendere tutta la nostra «roba»? In generale, forse, no, ma ci sono occasioni in cui sembra vero, psicologicamente. Per esempio, sebbene alcune persone si limitino semplicemente ad essere i possessori e i guidatori delle loro automobili, altri sono *automobilisti*; l'automobilista incallito preferisce *essere* un agente consumatore di benzina a quattro ruote piuttosto che un agente consumatore di cibo a due gambe, e il suo uso della prima persona tradisce questa identificazione:

Ho qualche problema nelle curve quando piove perché i miei pneumatici sono troppo consumati.

Così qualche volta allarghiamo i nostri confini; altre volte, in risposta a difficoltà reali o immaginarie, lasciamo che i nostri confini si restringano:

Non l'ho detto *io*! Non era il vero me che parlava. Sì, le parole sono uscite dalla mia bocca, ma rifiuto di riconoscerle come mie.

Ho ricordato questi discorsi familiari per attirare l'attenzione sulle analogie tra i nostri sé e i sé delle formiche e dei paguri, ma i discorsi attirano l'attenzione anche sulla differenza più importante: formiche e paguri non parlano. Il paguro è fatto in un modo tale che assicura a lui l'acquisizione di una conchiglia. La sua organizzazione, potremmo dire, *implica* una conchiglia, e quindi, in un senso *molto* debole, *rappresenta* tacitamente il paguro fornito di conchiglia, ma il paguro stesso non deve, in nessun senso più forte, *rappresentare se stesso* fornito di conchiglia. Non si dedica affatto all'auto-rappresentazione. A chi dovrebbe

rappresentare questo aspetto di se stesso e perché? Non ha bisogno di ricordare a se stesso questo aspetto della sua natura, poiché la sua costituzione innata si prende cura del problema, e non c'è nessun altro personaggio interessato nei paraggi. E le formiche e le termiti, come abbiamo notato, realizzano i loro progetti senza basarsi su nessuno schema progettuale o editto esplicitamente comunicato.

Noi, al contrario, siamo quasi costantemente impegnati a presentare noi stessi agli altri, e a noi stessi, e quindi a *rappresentare* noi stessi – tramite il linguaggio e i gesti, internamente ed esternamente. La differenza più ovvia del nostro ambiente che possa spiegare questa differenza del nostro comportamento è il comportamento stesso. Il nostro ambiente umano non contiene solo cibo e rifugi, nemici da combattere o evitare, cospecifici con cui accoppiarsi, ma parole, parole, parole. Queste parole sono elementi potenti del nostro ambiente che noi prontamente incorporiamo, ingeriamo, secerniamo, e che intrecciamo come ragnatele in sequenze *narrative* auto-protettive. In effetti, come abbiamo visto nel capitolo 7, quando diamo libero accesso a queste parole, questi veicoli di memi, esse tendono a prendere il sopravvento, a crearci, utilizzando il materiale grezzo che trovano nei nostri cervelli.

La *nostra* tattica fondamentale di auto-protezione, di auto-controllo e di auto-definizione non è quella di tessere ragnatele o costruire dighe, ma quella di raccontare storie, e più in particolare di architettare e controllare la storia che raccontiamo agli altri – e a noi stessi – su chi siamo. E proprio come i ragni non devono pensare, consciamente e deliberatamente, su come tessono le loro ragnatele; e come i castori, a differenza degli esseri umani ingegneri di professione, non pianificano consciamente e deliberatamente le strutture che costruiscono, noi (a differenza dei narratori umani *di professione*) non immaginiamo consciamente e deliberatamente quali narrazioni raccontare e come raccontarle. I nostri racconti vengono tessuti, ma per lo più noi non li tessiamo; essi ci tessono. La nostra coscienza umana – e la nostra individualità narrativa – è un loro prodotto, non la loro fonte.

È *come* se queste sequenze o flussi narrativi provenissero da un'unica fonte – non soltanto nell'ovvio senso fisico di fluire da una sola bocca, o da una sola penna o matita, ma in un senso più sottile: hanno l'effetto di incoraggiare qualunque pubblico a (tentare di) postulare un agente unitario a cui appartengono le parole, a cui esse si riferiscono; in breve, a postulare un *centro di*

gravità narrativa. I fisici apprezzano l'enorme semplificazione che ottieni quando postuli un centro di gravità in un oggetto, un unico punto rispetto al quale tutte le forze gravitazionali possono essere calcolate. Noi eterofenomenologi apprezziamo l'enorme semplificazione che ottieni quando postuli un centro di gravità narrativa in un corpo umano tessitore di narrazioni.

Come il sé biologico, questo sé psicologico o narrativo è ancora un'altra astrazione, non qualcosa nel cervello; ma tuttavia si rivela un potente e quasi tangibile attrattore di varie proprietà, è «il possessore ufficiale» di tutti quegli oggetti e caratteristiche che giacciono non reclamati. Chi possiede la tua macchina? Tu. Chi possiede i tuoi vestiti? Tu. Così, chi possiede il tuo corpo? Tu! Quando dici

Questo è il *mio* corpo

certamente nessuno pensa che tu voglia dire

Questo corpo possiede se stesso.

Ma cosa intendi dire, allora? Se ciò che dici non è né una bizzarra e inutile tautologia (questo corpo è auto-posseduto, o qualcosa del genere) né la pretesa che tu sia un'anima immateriale o un fantomatico burattinaio che possiede e guida questo corpo nello stesso modo in cui tu possiedi e guidi la tua automobile, cos'altro potresti intendere?

2. *Quanti sé per ogni cliente?*

Penso che potremmo capire meglio cosa significhi

Questo è il mio corpo,

se potessimo metterla a confronto con affermazioni alternative; che dire, per esempio, della seguente affermazione contrastante?

No, non è tuo; è *mio*, e non voglio condividerlo!

Se potessimo capire cosa proverebbero due (o più) sé a lottare per il controllo di un unico corpo, potremmo capire meglio che cos'è un singolo sé. Come scienziati del sé, vor-

remmo eseguire esperimenti controllati, in cui, variando le condizioni iniziali, potremmo capire che cosa esattamente deve avvenire, in che ordine e utilizzando quali risorse, affinché emerga un tale sé parlante. Ci sono condizioni nelle quali la vita va avanti ma non emerge alcun sé? Ci sono condizioni nelle quali emerge più di un sé? L'etica ci impedisce di eseguire tali esperimenti, ma, come molto spesso accade, noi possiamo servirci dei dati generati da qualche esperimento terribile che la natura stessa esegue e trarne cautamente le conclusioni.

In questo caso l'esperimento è la malattia mentale chiamata DPM (Disturbo della Personalità Multipla), in cui un singolo corpo umano *sembra* essere abitato da parecchi sé, ognuno, solitamente, con un nome proprio e un'autobiografia. Il concetto di DPM sembra a molte persone troppo strano e metafisicamente bizzarro per corrispondere a qualcosa di possibile – un fenomeno «paranormale» da scartare insieme all'ESP (percezione extrasensoriale), agli incontri ravvicinati del terzo tipo e alle streghe a cavallo di manici di scopa. Sospetto che alcune di queste persone abbiano fatto un semplice errore aritmetico: non si sono rese conto che due o tre o diciassette sé per corpo non è davvero metafisicamente più stravagante di un sé per corpo. Anche uno è troppo!

«Ho appena visto passare una macchina con cinque sé a bordo.»
«Cosa?? È pazzesco! Che razza d'assurdità metafisica è questa?»
«Sì, nella macchina c'erano anche cinque corpi.»
«Ah, be'; perché non hai detto così? In questo caso non ci sono problemi.»
«... o forse i corpi erano quattro, o tre – ma sicuramente cinque sé.»
«Cosa??!!»

Normalmente ci si accorda su un sé per ogni corpo, ma se un corpo ne può avere uno, perché in condizioni anormali non ne può avere di più?

Non intendo suggerire che non ci sia nulla di sciocccante o di profondamente misterioso riguardo al DPM. Tale fenomeno è, in effetti, di una stranezza incomparabile, non perché sfida le nostre presupposizioni su ciò che sia *metafisicamente* possibile, ma perché sfida le nostre presupposizioni su ciò che sia *umanamente* possibile, sui limiti della depravazione e della crudeltà

umana da una parte, e i limiti della creatività umana dall'altra. Giacché ci sono ora prove voluminose che non siamo di fronte a una manciata o a centinaia di casi di DPM, ma a migliaia di casi diagnosticati oggi, e tale malattia è quasi sempre dovuta ad abusi prolungati nella prima infanzia, di solito a sfondo sessuale e di una violenza rivoltante. Nicholas Humphrey ed io abbiamo indagato il DPM parecchi anni fa (Humphrey e Dennett, 1989), e abbiamo scoperto che si tratta di un fenomeno complesso che si estende ben al di là dei cervelli individuali di coloro che ne sono affetti.

Questi bambini sono cresciuti in circostanze così straordinariamente terrificanti e disorientanti che confesso di essere più stupito del fatto che essi siano riusciti a sopravvivere psicologicamente di quanto non lo sia nell'apprendere che questa conservazione di se stessi sia stata ottenuta tramite una disperata ridefinizione dei propri confini. Quello che fanno, quando sono messi di fronte a un conflitto e a un dolore soverchianti, è questo: «se ne vanno». Creano un confine che lascia fuori l'orrore, in modo che esso non capiti *a loro*; o non capita a nessuno o a qualche altro sé più capace di mantenere la sua organizzazione di fronte a tale assalto – almeno questo è quello che dicono di aver fatto, come meglio lo ricordano.

Come è possibile ciò? Che tipo di spiegazione potremmo dare, in ultima analisi a livello biologico, di un tale processo di divisione? Ci deve essere stato un unico ed intero sé che in qualche modo si è scisso come un'ameba? Come è possibile se il sé non è una parte fisica propria di un organismo o di un cervello, ma, come ho suggerito, un'astrazione? La risposta al trauma sembra, inoltre, così creativa che si è dapprincipio inclini a supporre che ci debba essere il lavoro di una specie di supervisore: un supervisore del programma cerebrale, un controllore centrale, o quello che sia. Ma dobbiamo ricordarci della colonia delle termiti, che sembrava anch'essa, a prima vista, richiedere un dirigente centrale a capo dell'organizzazione di un progetto così intelligente.

Noi ci siamo abituati a storie evoluzionistiche che iniziano da uno stato in cui un certo fenomeno non esiste ancora e terminano con uno stato in cui il fenomeno è decisamente presente. L'innovazione dell'agricoltura, dell'abbigliamento, delle abitazioni e degli strumenti, l'innovazione del linguaggio, l'innovazione della coscienza stessa, le prime innovazioni della vita sulla terra. Tutte queste storie possono essere raccontate. Ed ognuna di esse deve attraversare ciò che potremmo chia-

mare lo iato assolutistico. Esso è illustrato dalla curiosa argomentazione che segue (presa in prestito da Sanford, 1975):

Ogni mammifero ha un mammifero per madre,
ma c'è stato solo un numero finito di mammiferi, quindi
deve esserci stato un primo mammifero,
il che contraddice la prima premessa, quindi, contrariamente alle apparenze,
i mammiferi non esistono!

In questa argomentazione c'è qualcosa che deve essere indebolito. Cosa potrebbe essere? Il filosofo assolutistico o essenzialista è attratto dalle linee nette, le soglie, le «essenze», i «criteri». Per lui, deve esserci stato effettivamente un primo mammifero, un primo essere vivente, un primo momento della coscienza, un primo agente morale; ciò può essere identificato con qualunque prodotto della mutazione, qualunque candidato radicalmente nuovo, che soddisfi le condizioni essenziali – quali che siano – messe in luce dall'analisi.

Questa predilezione per i confini netti tra le specie ha rappresentato l'ostacolo intellettuale più grande che Darwin ha dovuto fronteggiare quando tentava di sviluppare la teoria dell'evoluzione (Richards, 1987). A questo modo di pensare si oppone una sorta di anti-essenzialismo che si trova a suo agio con i casi in penombra e con l'assenza di precise linee divisorie. Poiché i sé e le menti e perfino la coscienza stessa sono prodotti biologici (non elementi che si possono trovare nella tavola periodica della chimica), dovremmo aspettarci che le transizioni tra tali fenomeni e fenomeni differenti siano graduali, controverse, modificabili. Ciò non significa che ogni cosa sia sempre transitoria, graduale; le transizioni che sembrano graduali da una prospettiva ravvicinata sembrano di solito improvvise separazioni tra stabilizzazioni dell'equilibrio se viste da una posizione più lontana (Eldredge e Gould, 1972; ma anche Dawkins, 1982, pp. 101-109).

L'importanza di questo fatto per le teorie filosofiche (e le predilezioni dei filosofi) non è stata abbastanza ampiamente riconosciuta. Ci sono sempre state – e sempre ci saranno – alcune cose di passaggio, «anelli mancanti», quasi-mammiferi e simili che sfidano la definizione, ma il fatto è che *quasi* tutte le cose reali in natura (a differenza di quelle meramente possibili) tendono a collocarsi in grappoli o arcipelaghi di somiglianza separati nello spazio logico da enormi oceani vuoti. Non

abbiamo bisogno delle «essenze» o dei «criteri» per impedire al significato delle nostre parole di scivolare via da ogni parte; le nostre parole rimarranno ferme, e, come attratte da una forza di gravità, rimarranno abbastanza saldamente ancorate al più vicino arcipelago di somiglianza, anche se c'era – deve esserci stato – un breve istmo che una volta lo collegava tramite una serie di passi graduali a qualche altro arcipelago vicino. Quest'idea viene applicata tranquillamente a vari campi d'indagine. Ma molte persone che non hanno problemi ad adottare questo atteggiamento pragmatico nei confronti della notte e del giorno, degli esseri viventi e non viventi, dei mammiferi e premammiferi, diventano ansiose se vengono invitate ad adottare lo stesso atteggiamento pragmatico nei confronti dell'avere e non avere un sé. Esse pensano che in questo caso, come da nessun'altra parte in natura, deve esserci una divisione Tutto o Nulla e una distribuzione Uno per ogni Cliente.

La teoria della coscienza che abbiamo sviluppato scredita queste presupposizioni e il Disturbo della Personalità Multipla fornisce una buona illustrazione del modo in cui la teoria le mette alle strette. Le convinzioni che *non possono* esserci quasi-sé o sé sui generis, e che, inoltre, *deve* esserci un numero intero di sé associati con un corpo – ed è meglio che questo numero sia *uno*! – *non* sono auto-evidenti. Cioè, non sono più auto-evidenti, ora che abbiamo sviluppato in un certo dettaglio un'alternativa al Teatro Cartesiano con il suo Testimone o il suo Autore Centrale. Il DPM sfida queste presupposizioni da un versante, ma possiamo immaginare anche una sfida dall'altro versante: due o più corpi che condividono un unico sé! Un caso del genere potrebbe esserci effettivamente, in Inghilterra, a York: le gemelle Chaplin, Greta e Freda («Time», 6 Aprile 1981). Queste gemelle identiche, che hanno ora superato la quarantina e vivono in un pensionato, sembrano agire come un *unico* individuo; esse collaborano alla formulazione di un singolo atto linguistico, per esempio, finendo le frasi iniziate dall'altra senza difficoltà o parlando all'unisono, con solo mezzo secondo di ritardo. Per anni sono state inseparabili, tanto inseparabili quanto possono riuscire ad esserlo due gemelli che non sono siamesi. Alcune persone che hanno avuto a che fare con loro suggeriscono che la tattica più naturale ed efficace da adottare nei «loro» confronti è quella di pensare di avere di fronte *lei* piuttosto che *loro*.

La nostra posizione permette la possibilità teorica non solo del DPM ma anche del DPF (Disturbo della Personalità Frazio-

nata). È possibile? Perché no? Io non sto affatto suggerendo che queste gemelle siano collegate tramite la telepatia o ESP o qualunque altro tipo di legami occulti. Sto suggerendo solo che c'è una miriade di modi sottili e frequenti di comunicazione e coordinazione (tecniche, infatti, spesso altamente sviluppate da gemelli identici). Poiché queste gemelle hanno visto, udito, toccato, odorato e pensato cose molto simili in tutta la loro vita, e hanno iniziato con cervelli disposti, senza dubbio, a reagire in modo molto simile a tali stimoli, è possibile che non siano necessari vasti canali di comunicazione per mantenerle focalizzate su una qualche armonia elastica. (E comunque, quanto è unificato il più auto-posseduto tra noi?) Dovremmo esitare a prescrivere i limiti di una tale coordinazione pratica.

Ma in ogni caso, non ci sarebbero anche due sé individuali chiaramente definiti, uno per ogni gemella, e responsabili di mantenere in vita questa curiosa sciarada? Forse, ma se ognuna di queste donne fosse diventata così poco attenta a sé o così altruista (come diciamo) nella sua devozione alla causa comune, da aver più o meno trasceso se stessa (come diciamo anche) nella realizzazione del progetto? Come il poeta Paul Valéry disse una volta, in una deliziosa modificazione del motto del suo connazionale: «Talvolta io sono, talvolta io penso».

Nel capitolo 11 abbiamo visto che, sebbene appaia continua, la coscienza è in realtà lacunosa. Un sé può essere altrettanto lacunoso, può scivolare nel nulla così facilmente come una candela può essere spenta e poi riaccesa in un secondo momento, in circostanze più favorevoli. Tu sei la stessa persona le cui avventure infantili stai attualmente ricordando (talvolta chiaramente, talvolta confusamente)? Le avventure di quel bambino, il cui percorso nello spazio e nel tempo in apparenza ha accompagnato ininterrottamente quello del tuo corpo, sono le tue stesse avventure? Quel bambino con il tuo nome, quel bambino la cui firma scarabocchiata ti ricorda il modo in cui *tu* scrivevi il tuo nome, è (era) te quel bambino? Il filosofo Derek Parfit (1984) ha paragonato una persona ad un club, un tipo piuttosto differente di costruzione umana, che può cessare di esistere un anno ed essere ricostituito da alcuni dei suoi membri (precedenti?) qualche anno dopo. Sarebbe lo stesso club? Potrebbe esserlo, se, per esempio, il club avesse avuto una costituzione scritta che prevedesse esplicitamente tali interruzioni di esistenza. Ma non è detto. Potremmo conoscere tutti i fatti rilevanti nella situazione e capire tuttavia che non bastano a stabilire l'*identità* del (nuovo?) club. Per il punto di vista sui sé

– o le persone – che emerge qui, questa è l'analogia giusta; i sé non sono anime-perle che esistono indipendentemente, ma risultati dei processi sociali che ci creano e, come altri prodotti del genere, soggetti a improvvisi mutamenti di status. L'unico «momento» che ottimizza la traiettoria di un sé, o di un club, è la stabilità conferitagli dalla rete di credenze che lo costituisce, e quando queste credenze vengono a mancare esso viene a mancare, o permanentemente o temporaneamente.

È importante tenere presente ciò, quando si prende in esame un altro fenomeno molto amato e molto discusso dai filosofi, quello dei pazienti con il cervello diviso. Un cosiddetto cervello diviso è il risultato di una *commessurotomia*, un'operazione che recide il *corpus callosum*, l'ampio fascio di fibre che connette direttamente l'emisfero destro e sinistro della corteccia. Questo intervento lascia ancora agli emisferi una connessione indiretta, attraverso varie strutture del cervello mediano, ma è ovviamente un procedimento drastico, che deve essere eseguito solo quando non ci sono alternative. Serve ad alleviare alcune forme più gravi di epilessia che non sono altrimenti trattabili, impedendo alle scariche elettriche generate internamente e focalizzate originariamente in un emisfero di propagarsi anche nella parte opposta della corteccia. Secondo la leggenda filosofica standard, i pazienti con il cervello diviso possono essere «divisi in due sé» ma non soffrono per altri versi alcuna seria diminuzione delle loro capacità a seguito dell'operazione. La versione più accattivante di questa semplificazione afferma che i due «lati» originari della persona – l'emisfero sinistro più rigido e analitico, l'emisfero destro più rilassato, intuitivo e olistico – dopo l'operazione sono liberi di risplendere con più individualità, ora il normale e stretto gruppo di lavoro deve essere sostituito da una *détente* meno intima. Questa è una versione accattivante, ma è un'avventata esagerazione delle scoperte empiriche che la ispirano. In realtà, solo in una porzione molto limitata di casi sono stati osservati *alcuni* dei sintomi teoreticamente scioccanti della personalità multipla. (Vedi, per es., Kinsbourne, 1974; Kinsbourne e Smith, 1974; Levy e Trevarthen, 1976; Gazzaniga e LeDoux, 1978; Gazzaniga, 1985; Oakley, 1985; Dennett, 1985b).

Non deve sorprendere che i pazienti con il cervello diviso, come quelli con la visione cieca e la gente con Disturbi della Personalità Multipla, non riescano a soddisfare le aspettative filosofiche; non è, inoltre, colpa di nessuno. Non sono i filosofi (e molti altri interpreti, inclusi i ricercatori stessi) ad esagerare

deliberatamente le loro descrizioni dei fenomeni. È vero piuttosto che, nel loro sforzo di descrivere i fenomeni concisamente, si accorgono di essere attratti inesorabilmente, a causa delle risorse limitate del linguaggio ordinario, verso il modello semplicistico del Boss nel Corpo, dello Spettro nella Macchina, del Pubblico nel Teatro Cartesiano. Nicholas Humphrey ed io, mettendo a confronto le nostre accurate annotazioni su ciò che accadeva durante i vari incontri con persone affette da DPM, scoprimmo che molto spesso per descrivere quello che avevamo effettivamente visto cadevamo, malgrado noi stessi, in giri di frasi molto naturali ma seriamente fuorvianti. Thomas Nagel (1971), il primo filosofo a scrivere sui pazienti con il cervello diviso, offrì una spiegazione giudiziosa ed accurata dei fenomeni così come erano capiti allora, e, notando la difficoltà di formulare una spiegazione coerente, suppose che: «Forse ci è impossibile abbandonare certe maniere di concepire e rappresentare noi stessi, anche se sono scarsamente sorrette dalla ricerca scientifica» (1971, p. 397).

È effettivamente difficile, ma non impossibile. Il pessimismo di Nagel è esso stesso un'esagerazione. Non siamo appena riusciti, infatti, a liberarci dei modi tradizionali di pensare? Ora, alcune persone forse non vogliono abbandonare la visione tradizionale. Potrebbero perfino esserci delle buone ragioni – delle ragioni morali – per cercare di conservare il mito dei sé come perle nei cervelli, entità particolari, concrete ed enumerabili piuttosto che astrazioni, e per rifiutare la possibilità di quasi-sé, semi-sé, sé transitori. Ma questo è sicuramente il modo corretto per comprendere i fenomeni dei cervelli divisi. Per brevi periodi durante procedimenti sperimentali attentamente costruiti, alcuni di questi pazienti si biforcano nel rispondere a situazioni difficili, creando temporaneamente un secondo centro di gravità narrativa. Alcuni effetti di questa biforcazione possono sopravvivere indefinitamente in tracce di memoria mutuamente inaccessibili, ma a parte queste effettive tracce piuttosto primitive della biforcazione, la vita di un secondo sé rudimentale dura al massimo pochi minuti, un tempo insufficiente per accrescere quel tipo di autobiografia di cui è fatto un sé pienamente sviluppato. (Questo è vero in modo altrettanto ovvio anche per la maggior parte delle dozzine di sé frammentari sviluppati dai pazienti affetti da DPM; non ci sono, semplicemente, abbastanza ore di veglia in un giorno per consentire alla maggior parte di loro di immagazzinare più di pochi minuti di una biografia esclusiva a settimana.)

La separatezza delle narrative differenti è la linfa vitale dei sé differenti. Come il filosofo Ronald de Sousa (1976) nota:

Quando il Dottor Jekyll si trasforma in Mister Hyde, questa è una cosa strana e misteriosa. Sono due persone che si alternano nello stesso corpo? Ma ecco una cosa ancora più strana: anche il Dottor Juggle e il Dottor Boggle si alternano nello stesso corpo. Ma *essi sono così simili tra loro come due gemelli identici*! Tu protesti: perché dici allora che si sono trasformati uno nell'altro? E perché no? Se il Dottor Jekyll può trasformarsi in un uomo così differente come Hyde, sicuramente dovrà essere *più facile* per Juggle trasformarsi in Boggle, che è esattamente come lui.

Abbiamo bisogno di un conflitto o di una grande differenza per scuotere la nostra presupposizione naturale che ad ogni corpo corrisponda al massimo un agente. (p. 219)

Così *che cosa si prova* ad essere il sé dell'emisfero destro di un paziente dal cervello diviso? Questa è una domanda molto naturale[1] ed evoca un'immagine stupefacente – ed agghiacciante: eccoti qui, intrappolato nell'emisfero destro di un corpo il cui lato sinistro conosci intimamente (e ancora controlli) e il cui lato destro è ora così distante come il corpo di un estraneo. Vorresti dire al mondo che cosa si prova al tuo posto, ma non puoi! Sei tagliato fuori da tutte le comunicazioni verbali per la perdita delle tue linee telefoniche indirette con la stazione radio nell'emisfero sinistro. Tu fai del tuo meglio per segnalare la tua esistenza al mondo esterno, tormentando metà della tua faccia con smorfie e sorrisi asimmetrici, e talvolta (se sei un sé virtuoso dell'emisfero destro) scarabocchiando una parola o due con la tua mano sinistra.

Questo esercizio dell'immaginazione potrebbe proseguire in maniera ovvia, ma sappiamo che è una fantasia – una fantasia come le deliziose storie di Beatrix Potter su Peter Coniglio e i suoi amici, gli animali antropomorfi. Non perché «la coscienza risiede solo nell'emisfero sinistro» e non perché *non è possibile* il caso di qualcuno che si ritrovi in un tale pasticcio, ma semplicemente perché la commessurotomia *non* lascia ai singoli emisferi un'organizzazione vigile abbastanza distinta e robusta da sorreggere un sé separato.

Affermare che è «logicamente possibile» che ci sia un tale sé dell'emisfero destro in un paziente dal cervello diviso non costituisce una sfida per la mia teoria del sé; la mia teoria dice che non c'è e spiega perché: le condizioni per accumulare la sorta di ricchezza (e indipendenza) narrativa che costituisce un

sé pienamente sviluppato non sono presenti. La mia teoria è, similmente, non scalfita dalla pretesa – che io non mi sognerei di negare – che *possano* esserci conigli parlanti, ragni che scrivono messaggi in inglese nelle loro ragnatele e, per quanto ci riguarda, melanconici trenini a vapore. Potrebbero esserci, credo, ma non ci sono – così la mia teoria non deve spiegarli.

3. *L'insostenibile leggerezza dell'essere*

> Qualunque cosa accada, non importa quando o dove, siamo inclini a domandarci chi o che cosa ne sia responsabile. Questo ci porta a scoprire spiegazioni che altrimenti non riusciremmo a immaginare e ci aiuta a prevedere e a regolare non solo ciò che accade nel mondo, ma anche ciò che accade nella nostra mente. Ma se queste stesse tendenze ci spingessero a immaginare cose e cause che non esistono? In tal caso inventeremmo falsi dèi e superstizioni, e ne vedremmo la mano in tutte le coincidenze casuali. In realtà, forse, quella strana parola «io», come quando si dice «*Io ho avuto una buona idea*», riflette la stessa identica tendenza. Se siamo costretti a trovare una causa che causi tutto ciò che tu fai, ebbene, questo qualcosa ha bisogno di un nome. Tu lo chiami «*io*». Io lo chiamo «*tu*».
>
> Marvin Minsky (1985), p. 453

Secondo la mia teoria, un sé non è un punto matematico, ma un'astrazione definita dalle miriadi di attribuzioni e interpretazioni (incluse le auto-attribuzioni e le auto-interpretazioni) che hanno composto la biografia del corpo vivente di cui è il Centro di Gravità Narrativa. Come tale, svolge un ruolo singolarmente importante nell'incessante economia cognitiva di quel corpo vivente, perché, tra tutte le cose dell'ambiente di cui un corpo attivo deve farsi un modello mentale, nessuna è più cruciale di se stesso (vedi, per es., Johnson-Laird, 1988; Perlis, 1991).

Tanto per cominciare, ogni agente deve sapere che cosa *esso* è! Ciò può sembrare a prima vista o banale o impossibile. «Io sono me!» non è davvero molto informativo, e cos'altro si dovrebbe sapere – o si dovrebbe scoprire se non lo si sa già? Per gli organismi più semplici, in effetti, non c'è veramente un'autoconoscenza che va oltre la rudimentale sapienza biologica racchiusa in massime quali: «Quando hai fame, non mangiare te stesso!», «Quando c'è un dolore, è il tuo dolore!». In ogni

organismo, inclusi gli esseri umani, il riconoscimento di questi principi fondamentali di progettazione biologica è semplicemente «cablato» in lui – parte del progetto sottostante al suo sistema nervoso, come battere le palpebre quando qualcosa si avvicina all'occhio o tremare quando fa freddo. Un'aragosta potrebbe anche mangiare le chele di un'altra aragosta, ma la prospettiva di mangiare una delle sue chele è opportunatamente esclusa dalle sue possibilità. Le sue opzioni sono limitate, e quando essa «pensa» di muovere una chela, il suo «pensatore» è collegato in modo diretto e appropriato a quella chela che pensa di muovere. Gli esseri umani (e gli scimpanzé e forse qualche altra specie), d'altra parte, hanno più opzioni e quindi più fonti di confusione.

Qualche anno fa, le autorità del porto di New York hanno sperimentato un sistema radar comune utilizzabile dai proprietari di piccole barche. Un'unica antenna radar molto potente e piazzata a terra formava un'immagine radar del porto, che veniva trasmessa come un segnale televisivo ai proprietari delle barche che potevano risparmiare il costo di un radar installando un piccolo apparecchio televisivo nella loro barca. Di che utilità era questo sistema? Se ti fossi perso nella nebbia, bastava guardare nello schermo televisivo e avresti saputo che uno dei puntini luminosi in movimento sullo schermo eri tu – ma quale? Ecco un caso in cui la domanda «Quale cosa sono io?» non è né banale né irrimediabilmente senza risposta. Il dubbio svanisce di fronte a un semplice trucco: vira rapidamente, facendo compiere alla tua barca un giro completo; tu sei il puntino luminoso che traccia una piccola «O» sullo schermo – a meno che varie barche nella nebbia non stiano tentando di fare la stessa cosa nello stesso momento.

Questo metodo non è a prova d'errore, ma funziona la maggior parte delle volte, e illustra chiaramente un punto molto più generale: per controllare le varie e sofisticate attività in cui i corpi umani sono impegnati, il sistema di controllo del corpo (installato nel cervello) deve essere in grado di riconoscere un'ampia varietà di tipi differenti di segnali in ingresso come informazioni su se stesso, e quando sorgono delle perplessità o sopraggiunge lo scetticismo, l'unico modo affidabile (ma non a prova d'errore) per selezionare e attribuire correttamente l'informazione è di procedere a piccoli esperimenti: fai qualcosa e vedi cosa si muove.[2] Uno scimpanzé può imparare facilmente a raggiungere una banana attraverso un buco nella parete della sua gabbia, guidando i movimenti del suo braccio osservandolo

nello schermo di una televisione a circuito chiuso che riprende il suo braccio da una notevole distanza (Menzel e altri, 1985). Questa capacità di auto-riconoscimento è decisamente non banale, poiché dipende dalla capacità di notare la consonanza dei movimenti del braccio visti nello schermo con i movimenti del braccio non visti *ma voluti*. Cosa accadrebbe se gli sperimentatori inserissero un piccolo ritardo nella video registrazione? Quanto pensi *ti* ci vorrebbe per scoprire che stai osservando il *tuo* braccio (senza alcun suggerimento verbale dagli sperimentatori) se la trasmissione a circuito chiuso fosse differita, diciamo, di venti secondi?

Il bisogno di auto-conoscenza va oltre il problema dell'identificazione dei segnali esterni dei nostri movimenti corporei. Noi abbiamo bisogno di conoscere i nostri stati interiori, le nostre tendenze, le nostre decisioni, i nostri punti di forza e di debolezza; e il metodo fondamentale per ottenere questa conoscenza è essenzialmente lo stesso: fai qualcosa e «vedi» cosa si «muove». Un agente abbastanza evoluto deve sviluppare strategie per tenere nota sia della sua situazione corporea che di quella «mentale». Negli esseri umani, come abbiamo visto, queste strategie implicano soprattutto delle attività incessanti di racconto e controllo di storie, alcune fattuali e alcune fittizie. I bambini si esercitano a voce alta (si pensi a Snoopy, che dice a se stesso mentre sta seduto sul tetto della sua casetta: «Ecco l'asso della prima guerra mondiale...»). Noi adulti lo facciamo più elegantemente: prendiamo nota silenziosamente e senza sforzo delle differenze tra le nostre fantasie e le nostre narrazioni e riflessioni «serie». Il filosofo Kendall Walton (1973, 1978) e lo psicologo Nicholas Humphrey (1986) hanno mostrato, da prospettive differenti, l'importanza del dramma, del raccontare e del più fondamentale fenomeno del «far finta di essere...», nel fornire un allenamento agli esseri umani che si avviano al compito di tessere un sé.

Così noi costruiamo una storia definitoria su noi stessi, organizzata attorno a una specie di traccia radar basilare di auto-rappresentazione (Dennett, 1981a). La traccia non è il sé, naturalmente; è una *rappresentazione* di un sé (e la traccia di Ellis Island sullo schermo radar non è un'isola – è una rappresentazione di un'isola). Ciò che fa di una traccia la traccia del *me* e di un'altra traccia solo la traccia di un *lui*, di una *lei* o di un *esso*, non è come appare, ma ciò per cui è usata. Raccoglie e organizza le informazioni che *mi* riguar-

dano nello stesso modo in cui altre strutture nel cervello registrano le informazioni su Boston, o Reagan, o il gelato.

E dov'è la cosa a cui si *riferisce* la tua auto-rappresentazione? È ovunque tu sia (Dennett, 1978b). E *cos'è* questa cosa? È nulla di più, e nulla di meno, che il tuo centro di gravità narrativa.

Otto ribatte:

Il guaio dei centri di gravità è che non sono reali; sono delle finzioni teoriche.

Questo non è il guaio dei centri di gravità; è il loro vanto. Essi sono *magnifiche* finzioni, finzioni che ognuno sarebbe orgoglioso di aver creato. E i caratteri fantastici della letteratura sono ancora più meravigliosi. Si pensi ad Ismaele, in *Moby Dick*. «Chiamatemi Ismaele» è il modo in cui il testo inizia, e noi lo facciamo. Non chiamiamo il testo Ismaele, né chiamiamo Melville Ismaele. Chi o che cosa chiamiamo Ismaele? Chiamiamo Ismaele Ismaele, il meraviglioso personaggio fantastico che troviamo nelle pagine di *Moby Dick*. «Chiamatemi Dan», sentite dalle mie labbra, e voi lo fate, non chiamate le mie labbra o il mio corpo Dan, ma chiamate *me* Dan, la finzione teorica creata da ... beh, non da me ma dal mio cervello, agendo in concerto nel corso degli anni con i miei genitori e i miei parenti e i miei amici.

Questo va benissimo per te, ma *io* sono perfettamente reale. Posso essere stato creato dal processo sociale a cui hai appena alluso (devo esserlo stato, se non sono esistito prima della mia nascita), ma ciò che il processo ha creato è un *sé reale*, non un mero personaggio narrativo!

Penso di sapere a cosa vuoi arrivare. Se il sé non è una cosa reale, che succede alla responsabilità morale? Uno dei ruoli più importanti che viene attribuito al sé nel nostro schema concettuale tradizionale è quello di punto in cui lo scaricabarile finisce, come diceva il cartello di Harry Truman. Se i sé non sono reali – non sono *realmente* reali – la responsabilità non sarà, sempre e incessantemente, scaricata sulle spalle di qualcun altro? Se non c'è un ufficio presidenziale nel cervello, che ospita l'Autorità Superiore a cui tutte le decisioni possono essere attribuite, ci sembra di essere minacciati da una burocrazia kafkiana di homunculi, che, messi alle strette, si limitano a rispondere: «Non se la prenda con me, io sono solo un impie-

gato!». Il compito di costruire un sé che possa *prendersi* la responsabilità è un progetto sociale ed educativo importantissimo, ed è giusto essere preoccupati delle minacce alla sua integrità. Ma «una perla del cervello», un qualche cosa di reale «intrinsecamente responsabile», è un patetico ammennicolo che viene brandito come un amuleto di fronte a questa minaccia. L'unica speranza che abbiamo – e non è del tutto infondata – è quella di riuscire a capire, rimanendo all'interno della metodologia naturalistica, i modi in cui i cervelli producono le autorappresentazioni, aggiungendo così un sé responsabile all'equipaggiamento dei corpi che controllano, quando tutto va bene. Il libero arbitrio e la responsabilità morale sono sicuramente desiderabili, e, come ho provato a mostrare in *Elbow Room: The Varieties of Free Will Worth Wanting* (1984), si difendono meglio abbandonando il mito senza speranze e zeppo di contraddizioni di un'anima separata, distinta.

Ma io non esisto, allora?

Certo che esisti. Eccoti qua, seduto sulla tua sedia, mentre leggi il mio libro e avanzi delle obiezioni. E abbastanza curiosamente, la tua attuale incarnazione, sebbene sia una necessaria precondizione per la tua creazione, non è necessaria affinché la tua esistenza sia prolungata indefinitamente. Ora, se tu fossi un'anima, una perla di sostanza immateriale, potremmo «spiegare» la tua immortalità potenziale solo postulandola come una proprietà inesplicabile, una *virtus dormitiva* ineliminabile della sostanza mentale. E se tu fossi una perla di sostanza materiale, un gruppo del tutto speciale di atomi nel tuo cervello, la tua mortalità dipenderebbe dalle forze fisiche che li tengono insieme (potremmo chiedere ad un fisico qual è «il tempo di dimezzamento» di un sé). Se, invece, tu pensi te stesso come un centro di gravità narrativa, la tua esistenza dipende dal perdurare della narrazione (più o meno come *Le mille e una notte*), che potrebbe *in teoria* sopravvivere a infiniti cambiamenti del *mezzo*, essere teletrasportato così facilmente (in linea di principio) come il telegiornale della sera e immagazzinato indefinitamente come pura informazione. Se ciò che sei è questa organizzazione dell'informazione che ha strutturato il sistema di controllo del tuo corpo (o, per dirlo nel modo più usuale e provocatorio, se ciò che sei è il programma che gira nel tuo calcolatore cerebrale), allora potresti in linea di principio sopravvivere alla morte del tuo corpo, così come un programma può conservarsi

intatto anche dopo che sia stato distrutto il calcolatore sul quale è stato creato ed eseguito per la prima volta. Alcuni pensatori (per es., Penrose, 1989) ritengono che questa sia un'implicazione spaventosa e profondamente antiintuitiva della posizione che difendo. Ma se ambisci all'immortalità potenziale, le altre alternative sono semplicemente indifendibili.

14
LA COSCIENZA NELLA NOSTRA IMMAGINAZIONE

1. *Proviamo ad immaginare un robot cosciente*

I fenomeni della coscienza umana sono stati spiegati nei capitoli precedenti in termini di operazioni di una «macchina virtuale», una sorta di programma informatico evoluto (e evolventesi) che plasma le attività del cervello. Non esiste un Teatro Cartesiano; esistono solo Molteplici Versioni composte da processi di fissazione di contenuti che giocano vari ruoli semi-indipendenti nella più vasta economia tramite la quale il cervello controlla il viaggio del corpo umano attraverso la vita. La convinzione straordinariamente persistente che ci sia un Teatro Cartesiano è il risultato di una varietà di illusioni cognitive che sono state ora esposte e spiegate. I «qualia» sono stati sostituiti da complessi stati disposizionali del cervello e il sé (altrimenti noto come il Pubblico del Teatro Cartesiano, l'Autore Centrale o il Testimone) si rivela essere una valida astrazione, una finzione teorica piuttosto che un osservatore interno o un boss.

Se il sé è «soltanto» il Centro di Gravità Narrativa, e se tutti i fenomeni della coscienza umana sono «soltanto» i prodotti delle attività di una macchina virtuale realizzata nelle connessioni incredibilmente modificabili del cervello umano, allora, in linea di principio, un robot opportunamente «programmato», con un cervello costituito da un calcolatore a base di silicio, sarebbe cosciente, avrebbe un sé. Più precisamente, ci sarebbe un sé cosciente il cui corpo sarebbe il robot e il cui cervello sarebbe il calcolatore. Ad alcune persone questa implicazione della mia teoria sembra ovvia e ineccepibile. «*È naturale* che siamo macchine! Siamo solo macchine molto, molto complicate, macchine evolute e fatte di molecole organiche invece di metallo e silicio; inoltre siamo coscienti, quindi ci possono essere macchine coscienti – noi». Per tali lettori, questa implicazione rappresenta una conclusione a lungo attesa. Ciò che è stato

interessante per loro, spero, è la quantità di implicazioni non ovvie che abbiamo incontrato lungo il percorso, in particolare quelle che mostrano come una larga parte dell'immagine cartesiana di senso comune deve essere rimpiazzata quando impariamo di più sull'effettivo lavorio del cervello.

Altre persone, comunque, trovano così incredibile l'implicazione che ci possa essere, in linea di principio, un robot cosciente da farla apparire ai loro occhi una pura e semplice *reductio ad absurdum* della mia teoria. Una volta un mio amico reagì alla mia teoria asserendo sentitamente: «Ma, Dan, non riesco proprio ad immaginare un robot cosciente!». Alcuni lettori potrebbero essere portati ad approvare la sua affermazione. Essi dovrebbero resistere all'inclinazione però, perché egli si sbagliava. Il suo errore era semplice, ma utile perché attira l'attenzione su una confusione fondamentale che blocca i progressi negli sforzi di comprensione della coscienza. «Lo *sai* che è falso», risposi io, «hai spesso immaginato robot coscienti. Non è che non puoi immaginare un robot cosciente; è che non puoi immaginare *come* un robot possa essere cosciente».

Chiunque abbia visto R2D2 e C3PO in *Guerre stellari* o ascoltato Hal in *2001. Odissea nello spazio*, ha immaginato un robot cosciente (o un calcolatore cosciente – se il sistema è vivo e vegeto, come R2D2, o «costretto a letto», come Hal, non è davvero così importante per questo compito di immaginazione). È letteralmente un gioco da bambino immaginare il flusso di coscienza di una cosa «inanimata». Ed infatti i bambini lo fanno in continuazione. Non solo gli orsacchiotti di peluche, ma anche i trenini a vapore o altri oggetti dei cartoni animati hanno una vita interiore. Gli abeti stanno in silenzio nei boschi, temendo l'accetta del boscaiolo ma nello stesso tempo bramando di diventare alberi di Natale in case calde e accoglienti, circondati da allegri bambini. La letteratura per bambini (per non parlare della televisione) è piena zeppa di occasioni per immaginare la vita cosciente di queste mere cose. Gli artisti che illustrano queste fantasie di solito aiutano l'immaginazione dei bambini disegnando facce espressive su questi falsi agenti, ma ciò non è essenziale. Parlare – come fa Hal – servirà altrettanto bene, in assenza di una faccia espressiva, a produrre l'illusione che ci sia qualcuno lì, che si provi qualcosa ad essere Hal, o un orsacchiotto di peluche, o un trenino a vapore.

Ecco il problema, naturalmente: queste sono tutte illusioni – o così sembrano. Comunque, ci sono delle differenze tra loro. È ovvio che nessun orsacchiotto di peluche sia cosciente, ma non è

realmente ovvio che non lo possa essere nessun robot. Ciò che è ovvio è soltanto la difficoltà di immaginare *come* possa esserlo. Poiché ha trovato difficile immaginare *come* un robot possa essere cosciente, il mio amico è stato riluttante ad immaginare un robot che *fosse* cosciente – sebbene avrebbe potuto farlo facilmente. C'è una differenza enorme tra questi due sforzi d'immaginazione, ma la gente tende a confonderli. È effettivamente di una difficoltà strabiliante immaginare come il calcolatore/cervello di un robot possa sostenere la coscienza. Come *potrebbe* un complicato ammasso di eventi che elaborano informazioni su innumerevoli chip al silicio equivalere alle esperienze coscienti? Ma è altrettanto difficile immaginare come un cervello umano organico possa sorreggere la coscienza. Come *potrebbe* un complicato ammasso di interazioni elettrochimiche tra miliardi di neuroni equivalere alle esperienze coscienti? Eppure noi immaginiamo facilmente che gli esseri umani siano coscienti, sebbene non possiamo ancora immaginare come ciò sia possibile.

Come è possibile che il cervello sia la sede della coscienza? Questa è stata di solito considerata dai filosofi una domanda retorica, un invito a pensare che la sua risposta si collochi oltre la comprensione umana. Uno dei compiti principali di questo libro è stato quello di demolire tale presupposizione. Io ho sostenuto che *puoi* immaginare come tutto questo ammasso complicato di attività nel cervello equivalga all'esperienza cosciente. Il mio argomento è semplice: ti ho mostrato come farlo. Abbiamo visto che il modo migliore per farlo è quello di pensare al cervello come a una sorta di calcolatore. I concetti dell'informatica ci forniscono i supporti teorici necessari ad attraversare la *terra incognita* tra la nostra fenomenologia, come la conosciamo dall'introspezione, e i nostri cervelli, come ci sono rivelati dalla scienza. Pensando ai nostri cervelli come sistemi di elaborazione dell'informazione, possiamo diradare gradualmente la nebbia e mantenere la nostra rotta attraverso il grande spartiacque, scoprendo come sia possibile che i nostri cervelli producano tutti quei fenomeni. Ci sono molti perfidi trabocchetti da evitare – allettanti vicoli ciechi come l'Autore Centrale, «il riempire» e i «qualia», per esempio – e senza dubbio ci sono ancora imprecisioni residue e schietti errori nell'abbozzo che ho fornito, ma perlomeno ora possiamo vedere la struttura fondamentale del percorso.

Alcuni filosofi hanno dichiarato, comunque, che attraversare questo spartiacque è assolutamente impossibile. Thomas

Nagel (1974, 1986) ha sostenuto che dal livello oggettivo della fisiologia non è possibile arrivare al livello soggettivo della fenomenologia. Più recentemente Colin McGinn ha sostenuto che la coscienza ha una «struttura nascosta» che giace al di là sia della fenomenologia che della fisiologia e che sebbene questa struttura nascosta possa colmare la lacuna, ci è probabilmente inaccessibile per sempre.

Il tipo di struttura nascosta che immagino non si situerebbe a nessuno dei livelli proposti da Nagel: sarebbe collocata tra essi. Né fenomenologico né fisico, questo livello mediante non sarebbe (per definizione) forgiato sul modello di nessuna delle due parti dello spartiacque e quindi non si scoprirebbe incapace di raggiungere la parte opposta. La sua caratterizzazione richiederebbe una radicale innovazione concettuale (che è probabilmente fuori della nostra portata). (McGinn, 1991, pp. 102-103.)

Il livello di descrizione del «software» o della «macchina virtuale» che ho sfruttato in questo libro è proprio quel tipo di livello capace di mediazione di cui parla McGinn; non è esplicitamente fisiologico o meccanico eppure è capace di fornire i collegamenti necessari con il meccanismo cerebrale da una parte, mentre d'altra parte non è esplicitamente fenomenologico eppure è capace di fornire i necessari collegamenti con il mondo del contenuto, i mondi (etero-)fenomenologici. Ci siamo riusciti! *Abbiamo* immaginato come un cervello possa produrre l'esperienza cosciente. Perché McGinn pensa che sia al di fuori della nostra portata impegnarci in questa «radicale innovazione concettuale»? Egli ha forse sottoposto a un'analisi rigorosa e dettagliata i vari approcci alla mente che utilizzano il livello software e dimostrato la loro futilità? No; non li esamina affatto. Egli non prova nemmeno ad immaginare il livello intermedio che postula; non fa altro che notare che gli sembra ovvio che non c'è nulla da sperare a questo riguardo.

Questa «ovvietà» spuria costituisce un grande ostacolo al progresso di comprensione della coscienza. È la cosa più naturale del mondo pensare che la coscienza avvenga in una specie di Teatro Cartesiano e supporre che non ci sia davvero nulla di sbagliato in questo modo di pensare. Ciò sembra ovvio fino a che non si osservi più attentamente quello che possiamo imparare sulle attività del cervello e si inizi a provare ad immaginare, in dettaglio, un modello alternativo. Allora succede più o meno la stessa cosa che accade quando impariamo come un

mago sul palcoscenico effettua un gioco di prestigio. Quando osserviamo attentamente da dietro le quinte, scopriamo che non avevamo visto in realtà quello che pensavamo di aver visto dalla platea. L'ampia lacuna tra fenomenologia e fisiologia si restringe un po'; vediamo che alcune delle «ovvie» caratteristiche della fenomenologia non sono affatto reali: non c'è riempimento con figmento; non ci sono qualia intrinseci; non c'è una fonte centrale del significato e dell'azione; non c'è un luogo magico dove la comprensione avviene. Infatti, non c'è Teatro Cartesiano; la stessa distinzione tra esperienze sul palcoscenico e processi dietro le quinte perde le sue attrattive. Noi abbiamo àncora un'abbondanza di fenomeni stupefacenti da spiegare, ma alcuni dei più sconcertanti effetti speciali semplicemente non esistono affatto, e quindi non richiedono spiegazione.

Una volta fatto qualche progresso nel compito difficile – immaginare *come* un cervello produca i fenomeni della coscienza – finiamo col fare qualche leggera modificazione nel compito facile: immaginare qualcuno o qualcosa che sia cosciente. Possiamo continuare a pensare che il modo migliore di immaginarlo sia postulando una specie di flusso di coscienza, ma non dotiamo più questo flusso di coscienza di tutte le sue proprietà tradizionali. Ora che il flusso di coscienza è stato riconcepito come le operazioni di una macchina virtuale realizzata nel cervello, non è più «ovvio» che stiamo cedendo ad un'illusione quando immaginiamo che un tale flusso avvenga in un cervello/calcolatore di un robot, per esempio.

McGinn invita i suoi lettori ad arrendersi insieme a lui: è semplicemente impossibile immaginare come il software potrebbe rendere cosciente un robot. Non provate neanche, egli dice. Altri filosofi hanno incoraggiato questa tendenza escogitando esperimenti ideali che «funzionano» proprio perché dissuadono il lettore dal provare ad immaginare, in dettaglio, come il software potrebbe essere capace di ciò. È abbastanza curioso che i due più noti implichino entrambi un riferimento alla Cina: la Nazione Cinese di Ned Block (1978) e la Stanza Cinese di John Searle (1980, 1982, 1984, 1988).[1] Si basano entrambi sullo stesso modo di fuorviare l'immaginazione, e poiché l'esperimento ideale di Searle è stato quello più discusso, io concentrerò su di esso la mia attenzione. Searle ci invita ad immaginare lui stesso chiuso in una stanza, mentre simula manualmente un gigantesco programma di IA che appare in grado di capire il cinese. Egli stabilisce che il programma supera il test di Turing, frustrando tutti i tentativi fatti dagli

interlocutori umani per distinguerlo da una persona che autenticamente comprende il cinese. Da questa indistinguibilità meramente comportamentale non discende, egli afferma, che ci sia un'autentica comprensione del cinese, o qualsiasi coscienza cinese, nella Stanza Cinese. Searle, chiuso nella stanza a manipolare affannosamente le stringhe di simboli del programma seguendo il listato, non acquista facendo ciò alcuna comprensione del cinese, né c'è qualcun altro nella stanza che possa capire il cinese (questo è «semplicemente ovvio», come direbbe Frank Jackson).

Questo esperimento ideale dovrebbe dimostrare l'impossibilità di ciò che Searle chiama la «IA forte», la tesi che un «calcolatore numerico adeguatamente programmato con i giusti input e output avrebbe perciò una mente nello stesso senso esatto in cui si dice che gli esseri umani hanno una mente» (Searle, 1988a). Negli ultimi dieci anni, ci sono state innumerevoli reazioni alle molte versioni offerte da Searle di questo esperimento ideale e sebbene filosofi e non filosofi abbiano sempre trovato alcune pecche nel suo esperimento ideale inteso come argomento logico,[2] è innegabile che la sua «conclusione» continui a sembrare «ovvia» a molte persone. Perché? Perché molte persone non immaginano effettivamente l'esempio nei dettagli necessari.

Ecco un esperimento informale che ci aiuterà a capire se la mia diagnosi è corretta. Come prima cosa, immaginiamoci un breve estratto dal dialogo vincente della Stanza Cinese con il giudice del test di Turing. (Per comodità, è stato tradotto dal cinese all'italiano.)

GIUDICE: La sai quella dell'irlandese che ha trovato la lampada magica? Dopo che l'ha strofinata è apparso un genio che gli ha concesso tre desideri. «Mi piacerebbe un boccale di Guinness!», disse l'irlandese e immediatamente fu accontentato. L'irlandese iniziò a trangugiare avidamente, ma il livello della birra nel bicchiere si ristabiliva magicamente. Dopo un po' il genio cominciò a spazientirsi. «Allora, il tuo secondo desiderio?», chiese. Tra un sorso e l'altro l'irlandese replicò, «Beh, credo proprio che ne prenderò un altro!».
STANZA CINESE: Molto divertente. No, non la sapevo – ma, sai, trovo che le barzellette sui gruppi etnici siano di cattivo gusto. Ho riso, mio malgrado, ma credo proprio che dovresti trovare un altro argomento di conversazione.

G: Hai ragione, ma ti ho raccontato la barzelletta perché vorrei che me la spiegassi.
SC: Che noia! Non si dovrebbero mai spiegare le barzellette.
G: Tuttavia, questa è la mia domanda del test. Puoi spiegarmi come e perché la barzelletta «funziona»?
SC: Se insisti. Vedi, si basa sul presupposto che il magico riempimento continuo del bicchiere proseguirà per sempre, cosicché l'irlandese avrà tutta la birra che vorrà. Egli non ha quindi nessuna ragione per volere un duplicato, ma è così sciocco (questa è la parte che mi dà fastidio) o così istupidito dall'alcool che non lo capisce, e quindi, estasiato per la realizzazione del primo desiderio, chiede un secondo boccale di birra. Questi presupposti di fondo non sono veri, naturalmente, fanno semplicemente parte dell'ambientazione tradizionale che si crea quando vengono raccontate le barzellette: in questi momenti noi sospendiamo la nostra incredulità nei confronti della magia, e così via. Tra l'altro, potremmo immaginare una continuazione un po' complicata che dimostrerebbe che l'irlandese, dopo tutto, non si era sbagliato nel chiedere il suo secondo desiderio – forse stava pensando di dare una festa e un unico bicchiere non si sarebbe riempito abbastanza velocemente per spegnere la sete di tutti i suoi ospiti (e non serve a nulla accumulare la birra in anticipo – sappiamo tutti come la birra vecchia perda il suo gusto). Noi di solito non pensiamo a tali complicazioni, il che spiega in parte perché le barzellette funzionano. Ti basta?

La conversazione non è brillantissima, ma supponiamo che sia abbastanza buona da ingannare il giudice. Ora, noi siamo invitati ad immaginare che tutti questi discorsi pronunciati dalla SC siano prodotti da un programma gigantesco che Searle con diligenza simula manualmente. È difficile da immaginare? Naturalmente, ma poiché Searle stabilisce che il programma supera il test di Turing e poiché questo livello di sofisticazione della conversazione rientrerebbe sicuramente tra i suoi poteri, noi non seguiremmo le indicazioni se non provassimo ad immaginare le complessità di un programma capace di generare una conversazione del genere. Naturalmente dovremmo anche immaginare che Searle non abbia la minima idea di ciò che sta facendo nella Stanza Cinese; egli ha a che fare solo con stringhe di zero e uno che manipola secondo il programma. È importante notare, tra l'altro, che Searle ci invita ad immaginare che gli oggetti della sua attenzione siano degli impenetrabili carat-

teri cinesi invece che zeri e uno, perché questo ci può attrarre verso la (ingiustificata) supposizione che il programma gigantesco possa lavorare «accoppiando» semplicemente in qualche modo i caratteri cinesi in ingresso con altri caratteri cinesi in uscita. Nessun programma del genere funzionerebbe, naturalmente – i discorsi in italiano della SC si «accoppiano» alle domande del giudice?

Un programma che potesse davvero generare le frasi pronunciate dalla SC in risposta alle domande del Giudice sembrerebbe qualcosa del genere, se visto in azione, non dal livello limitato di Searle, ma dal livello della macchina virtuale. Non appena vengono analizzate le prime parole, «La sai quella ...», alcuni folletti del programma incaricati di rilevare le barzellette si metterebbero in attività, chiamando in causa varie strategie per trattare con racconti fantastici, espressioni con «secondi fini» e simili, cosicché quando le parole «lampada magica» arriverebbero ad essere analizzate, il programma avrebbe già assegnato una bassa priorità alle risposte che puntualizzano la non esistenza di cose come le lampade magiche. Un insieme molto variegato di *frames* narrativi (Minsky, 1975) o *scripts* (Schank e Abelson, 1977) per barzellette standard sulla magia dei geni verrebbero attivati, creando varie aspettative per la prosecuzione, che sarebbero, però, fatte saltare dalla battuta conclusiva, che implica uno *script* più banale (quello della richiesta identica), di cui il programma non avrebbe mancato di rilevare la comparsa inaspettata.... Nello stesso tempo, sarebbero stati messi in allarme anche dei folletti sensibili alle connotazioni negative delle barzellette sui gruppi etnici, che condurrebbero poi al secondo tema della prima risposta della SC... E così via, in modo molto più dettagliato di quello che ho provato ad abbozzare ora.

Qualunque programma che possa davvero sostenere una conversazione simile a quella mostrata dovrebbe avere un sistema estremamente flessibile, sofisticato e stratificato, stracolmo di «conoscenze sul mondo» e meta-conoscenze e meta-meta-conoscenze sulle proprie risposte, le probabili risposte del suo interlocutore, e molte, molte altre cose. Searle non nega, naturalmente, che un programma possa avere una struttura così complicata. Egli cerca semplicemente di ostacolare la nostra immaginazione. Ma se dobbiamo immaginare correttamente la situazione, non siamo solo autorizzati ma obbligati ad immaginare che il programma che Searle sta simulando manualmente abbia questa struttura – ed altro, se solo potes-

simo immaginarlo. Ma allora non è più *ovvio* che non siamo di fronte ad una autentica comprensione della barzelletta. *Forse* i miliardi di azioni compiute da tutte quelle parti altamente strutturate produrrebbero una autentica comprensione nel sistema, dopo tutto. Se la tua risposta a questa ipotesi è che non hai la più pallida idea sulla presenza o l'assenza di una comprensione autentica in un sistema così complesso, allora ciò è già abbastanza per mostrare che l'esperimento ideale di Searle dipende, illecitamente, dal fatto che ci limitiamo ad immaginare un caso troppo semplice, un caso irrilevante, e da esso traiamo le conclusioni «ovvie».

Ecco come le nostre intuizioni vengono fuorviate. Noi capiamo abbastanza chiaramente che se ci fosse una reale comprensione in un tale sistema gigantesco, non sarebbe la comprensione di Searle (poiché egli è solo un ingranaggio nel meccanismo, ignaro del contesto in cui agisce). Inoltre, capiamo chiaramente che non c'è nulla che assomigli sia pur lontanamente alla comprensione autentica in qualunque blocco del programma abbastanza piccolo da poter essere immaginato facilmente – qualunque cosa esso sia, è solo un'inconsapevole routine per trasformare stringhe di simboli in altre stringhe di simboli secondo qualche ricetta meccanica o sintattica. A questo punto interviene la premessa nascosta: sicuramente *un incremento degli stessi ingredienti*, per quanto massiccio possa essere, non condurrà mai ad una comprensione autentica. Ma perché mai si dovrebbe pensare così? I dualisti cartesiani lo penserebbero, perché pensano che anche i cervelli umani lasciati a se stessi non raggiungerebbero la comprensione; secondo Cartesio, è necessaria l'anima immateriale per produrre il miracolo della comprensione. Se, d'altra parte, siamo materialisti convinti che in un modo o nell'altro i nostri cervelli siano responsabili essi stessi, senza assistenza miracolosa, della nostra comprensione, allora dobbiamo ammettere che la comprensione autentica sia in qualche modo raggiunta tramite un processo composto da interazioni tra una schiera di sottosistemi dei quali nessuno comprende nulla di per sé. L'argomentazione che inizia «questa piccola porzione di attività cerebrale non comprende il cinese, né lo comprende questa porzione più grande che la ingloba...» è destinata a dimostrare l'indesiderata conclusione che anche l'attività dell'intero cervello è insufficiente a spiegare la comprensione del cinese. È *difficile immaginare come* «un incremento degli stessi ingredienti» possa portare alla comprensione, ma abbiamo buone ragioni per credere che

sia così, quindi, in questo caso, dovremmo sforzarci di più, non abbandonare.

Come potremmo sforzarci di più? Con l'aiuto di alcuni utili concetti: i concetti del livello intermedio del software che sono stati forgiati dagli informatici proprio per aiutarci a seguire le complessità altrimenti inimmaginabili dei vasti sistemi. Ai livelli intermedi noi vediamo molte entità che sono più o meno invisibili ai livelli più microscopici, come i «folletti» menzionati prima, a cui una briciola di quasi-comprensione è stata attribuita. Allora non diventa difficile immaginare come «un incremento degli stessi ingredienti» possa equivalere all'autentica comprensione. Tutti questi folletti ed altre entità sono ordinati in enormi sistemi, le cui attività si organizzano attorno al proprio Centro di Gravità Narrativa. Searle, che si affanna nella Stanza Cinese, non comprende il cinese, ma egli non è solo nella stanza. C'è anche il Sistema, la SC, ed è a *questo* sé che dovremmo attribuire qualunque comprensione della barzelletta.

Questa risposta al suo esempio viene chiamata da Searle la Risposta dei Sistemi. È stata la risposta tipica degli addetti ai lavori dell'IA sin dalle prime apparizioni di questo esperimento ideale, più di una decina di anni fa, ma viene raramente apprezzata dalle persone al di fuori del campo dell'IA. Perché? Probabilmente perché non hanno imparato come immaginare un sistema del genere. Non riescono semplicemente ad immaginare come la comprensione possa essere una proprietà che emerge da una molteplicità di quasi-comprensioni distribuite in un grande sistema. Questa gente sicuramente non può riuscire a immaginarlo se non prova, ma come possiamo fare per aiutarla in questo difficile esercizio? Pensare al software come composto di homunculi che quasi-comprendono significa «barare» o è proprio il sostegno giusto con cui possiamo aiutare l'immaginazione ad orientarsi in complessità astronomiche? Searle commette una petizione di principio. Ci invita ad immaginare che il programma gigantesco abbia un'architettura che consiste in semplici scorse di elenchi e accoppiamenti diretti di alcune stringhe di caratteri cinesi ad altre, come se un programma del genere potesse stare, adeguatamente, per qualunque altro programma. Non serve a nulla immaginare un programma così semplice e assumere che *esso* sia il programma che Searle sta simulando, poiché nessun programma del genere potrebbe produrre il tipo di risultato che supererebbe il test di Turing, come re-

clamizzato. (Per un'argomentazione simile e la sua confutazione, vedi Block, 1982; e Dennett, 1985).

La complessità è importante: se non lo fosse, ci sarebbe un'argomentazione molto più breve contro l'IA forte: «Ehi, vedi questa calcolatrice portatile; non capisce il cinese, e qualunque calcolatore si possa immaginare è solo una gigantesca calcolatrice portatile; quindi nessun calcolatore potrà mai capire il cinese. C.V.D.». Quando mettiamo in conto la complessità, come dobbiamo fare in questo caso, dobbiamo davvero metterla in conto – e non solo far finta di metterla in conto. Potrà essere difficile, ma fintanto che non lo facciamo, non dobbiamo fidarci di qualsiasi nostra intuizione su ciò che è «ovviamente» assente. Come nel caso di Mary, la scienziata del colore, presentato da Frank Jackson, l'esperimento ideale di Searle genera una convinzione chiara e forte solo quando non ci curiamo di seguire le indicazioni. Queste pompe di intuizione sono difettose; non accrescono la nostra immaginazione, la fuorviano.

Ma che dire, allora, delle mie pompe di intuizione? Che dire, per esempio, del robot Shakey o del CADBLIND Mark II o del paziente affetto da visione cieca e allenato con un biofeedback? Non sono ugualmente sospette, ugualmente colpevoli di fuorviare il lettore? Ho certamente fatto del mio meglio nel raccontare queste storie in un modo che mi permettesse di guidare la tua immaginazione lungo determinati percorsi e evitasse di farti impantanare in complessità che non giudicavo necessarie al punto che stavo cercando di chiarire. C'è comunque una certa asimmetria: le mie pompe di intuizione sono usate, nella maggior parte dei casi, per aiutarti ad immaginare nuove possibilità, non per convincerti che certe prospettive sono impossibili. Ci sono delle eccezioni. Le mie variazioni sul tema del cervello nella vasca con cui ho aperto il libro dovevano servire a convincerti dell'impossibilità di certi tipi di inganni; e alcuni degli esperimenti ideali del capitolo 5 dovevano far vedere che, senza un Teatro Cartesiano, *sarebbe impossibile* distinguere le revisioni del contenuto orwelliane da quelle staliniane. Questi esperimenti ideali, comunque, procedevano accentuando l'intensità delle «opposizioni»; gli esempi della donna con il cappello alla festa e della donna con i capelli lunghi e gli occhiali, per esempio, dovevano servire ad affinare la stessa intuizione che ho poi cercato di screditare con un'argomentazione.

In ogni modo, il lettore faccia attenzione: le mie pompe di intuizione, come quelle di chiunque altro, non sono dimostra-

zioni dirette come potrebbero sembrare; sono più arte che scienza. (Per ulteriori avvertimenti riguardo agli esperimenti ideali dei filosofi, vedi Wilkes, 1988.) Se ci aiutano a concepire nuove possibilità, che possiamo poi confermare con metodi più sistematici, sono un bene; se ci adescano sul fiorito sentiero della facilità, sono un danno. Anche i buoni utensili possono essere usati erroneamente, e come tutti gli altri lavoratori, procederemo meglio comprendendo come funzionano i nostri strumenti.

2. *Cosa si prova ad essere un pipistrello*

L'esperimento ideale sulla coscienza più autorevole e più citato è «Che cosa si prova ad essere un pipistrello?» di Thomas Nagel (1974). Egli risponde a tale domanda affermando che ci è impossibile immaginarlo. Evidentemente questa tesi è congeniale a molte persone; e talvolta si vede il suo scritto citato da scienziati come se fosse – rarità delle rarità – un «risultato» filosofico, una dimostrazione accettata di un fatto che qualsiasi teoria è obbligata a rispettare.

Nagel sceglie molto bene le creature da prendere in esame. I pipistrelli, essendo mammiferi, sono abbastanza simili a noi da giustificare la convinzione che siano *ovviamente* coscienti. (Se egli avesse scritto «Che cosa si prova ad essere un ragno?», molti sarebbero stati propensi a chiedere perché mai era così sicuro che si dovesse provare qualcosa ad essere un ragno.) D'altra parte, grazie al loro sistema di ecolocazione – i pipistrelli possono «vedere con le orecchie» – essi sono anche abbastanza differenti da noi da farci avvertire un profondo divario. Se egli avesse scritto un saggio intitolato «Che cosa si prova ad essere uno scimpanzé?» o, più a proposito, «Che cosa si prova ad essere un gatto?», la sua pessimistica conclusione non sarebbe più stata ritenuta quasi unanimemente *ovvia*. Ci sono molte persone che sono fortemente convinte di sapere *esattamente* che cosa si prova ad essere un gatto. (Si sbagliano, naturalmente, a meno che non abbiano integrato tutta la loro empatica e amorevole osservazione con un'enorme quantità di ricerche fisiologiche, ma errerebbero dalla parte sbagliata, secondo Nagel.)

In ogni caso, sembra che la maggior parte della gente accetti di buon grado il «risultato» di Nagel sulla nostra incapacità di accedere alla coscienza dei pipistrelli. Alcuni filosofi, comun-

que, lo hanno messo in discussione, e per buone ragioni (Hofstadter, 1981; Hardin, 1988; Leiber, 1988; Akins, 1990). Innanzitutto dobbiamo fare chiarezza su cosa afferma. Non è la tesi epistemologica o probativa che dice che se qualcuno riuscisse («per caso») ad immaginare che cosa si prova ad essere un pipistrello, noi non saremmo mai in grado di confermare che tale successo immaginativo si sia verificato. Afferma piuttosto che noi esseri umani non abbiamo e non potremo mai acquisire il «conquibus», il meccanismo rappresentazionale, necessario per rappresentare a noi stessi che cosa si prova ad essere un pipistrello.

La distinzione è importante. Nel capitolo 12 ci siamo occupati di uno sforzo di immaginazione abbastanza simile: che cosa si doveva provare ad essere un abitante di Lipsia che ascolta le cantate di Bach per la prima volta. Il problema epistemologico è difficile, ma può essere affrontato direttamente dalle ricerche usuali. Immaginare quei tipi di esperienza, e le differenze con le nostre esperienze di Bach, è un problema che riguarda l'indagine storica, culturale, psicologica e, forse, fisiologica. Alcune di queste cose possono essere immaginate abbastanza facilmente, incluse alcune delle differenze più notevoli con le nostre esperienze, ma se dovessimo tentare di metterci nella stessa sequenza di stati esperienziali di cui una tale persona avrebbe goduto, ci troveremmo di fronte a difficoltà ben maggiori. Il compito ci obbligherebbe a subire vaste trasformazioni – dimenticare molto di quello che sappiamo, perdere delle associazioni e delle abitudini, acquistare nuove abitudini e associazioni. Possiamo sfruttare le nostre ricerche «in terza persona» per elencare le necessarie trasformazioni, ma intraprenderle effettivamente implicherebbe dei costi terribili in termini di isolamento dalla nostra cultura contemporanea – ci sarebbe proibito ascoltare la radio, leggere qualsiasi notizia sullo sviluppo politico e sociale successivo a Bach, e così via. Non è necessario percorrere tutto questo cammino per imparare qualcosa sulla coscienza degli abitanti di Lipsia del XVIII secolo.

Le stesse osservazioni si possono fare sul compito di immaginare che cosa si prova ad essere un pipistrello. Noi dovremmo essere interessati a ciò che possiamo conoscere sulla (eventuale) coscienza del pipistrello e non alla possibilità o meno di trasformare la nostra mente temporaneamente o permanentemente in una mente di pipistrello. Nel capitolo 12, abbiamo scalzato la presupposizione che ci siano delle proprietà «intrinseche» – i qualia – che costituiscono *ciò che si prova* quando si ha una

qualche esperienza cosciente, e, come Akins (1990) sottolinea, anche se ci fossero proprietà residue non disposizionali, non relazionali delle esperienze dei pipistrelli, arrivare ad avere una conoscenza intima di esse pur rimanendo all'oscuro dei fatti che la ricerca ci può dischiudere sulla struttura sistematica della percezione e del comportamento dei pipistrelli, ci lascerebbe nell'ignoranza rispetto a cosa si prova ad essere un pipistrello. C'è sicuramente una quantità enorme di fatti che possiamo venire a sapere su cosa si prova ad essere un pipistrello e né Nagel né nessun altro ci ha fornito delle buone ragioni per credere che ci sia qualcosa di interessante e teoreticamente importante che ci è precluso.

Nagel sostiene che qualsiasi ampliamento della conoscenza in terza persona non potrà mai dirci che cosa si prova ad essere un pipistrello; nego decisamente la sua tesi. Come si potrebbe risolvere questa disputa? Impegnandosi in qualcosa che inizia come un gioco da bambini – un gioco in cui una persona immagina che cosa si prova ad essere *x*, e l'altro poi prova a dimostrare che c'è qualcosa di sbagliato in quel particolare episodio di eterofenomenologia.
Ecco alcuni esempi di riscaldamento:

A: Ecco Pooh, l'orsacchiotto di peluche, mentre pensa quanto sarebbe bello mangiare del miele a colazione!
B: Sbagliato. Gli orsacchiotti di peluche non hanno il modo di distinguere il miele dalle altre cose. Non hanno nessun organo di senso in funzione, e neppure uno stomaco. Gli orsacchiotti di peluche sono pieni di imbottitura inerte. Non si prova nulla ad essere un orsacchiotto di peluche.
A: Ecco Bambi, il daino, mentre ammira un meraviglioso tramonto, fino a che quel bel cielo arancione non gli ricorda il giubbotto del cacciatore cattivo!
B: Sbagliato. I daini non vedono i colori (o meglio, possono avere solo una specie di visione dicromatica). Se anche possono essere coscienti, sicuramente non distinguono colori come l'arancione.
A: Ecco Peppe, il pipistrello, mentre percepisce con il suo sonar che l'oggetto volante in picchiata verso di lui non è suo cugino Roberto, ma un'aquila, con le ali spiegate e gli artigli pronti ad uccidere!
B: Aspetta un attimo – quanto hai detto che era distante l'aquila? L'ecolocazione dei pipistrelli funziona solo in un raggio di pochi metri.

A: Uhm... beh, comunque l'aquila ormai era distante solo due metri!
B: Ah, allora le cose sono più complicate. Quali sono i limiti della capacità di risoluzione dell'ecolocazione dei pipistrelli? Viene usato per identificare gli oggetti o solo come un allarme e un perlustratore per la cattura? Un pipistrello sarebbe in grado di distinguere artigli chiusi da artigli aperti usando semplicemente l'ecolocazione? Lo dubito, ma per saperne di più dovremmo escogitare nuovi esperimenti; tra l'altro, avremmo bisogno, ovviamente, anche di altri esperimenti per scoprire se i pipistrelli sono capaci di rilevare e reidentificare i loro consanguinei. Alcuni mammiferi possono farlo, ma abbiamo buone ragioni per credere che altri mammiferi non sappiano nulla di tali faccende.

I tipi di indagine suggeriti da questo esercizio ci porterebbero molto addentro ad una spiegazione della struttura del mondo percettivo e comportamentale dei pipistrelli, di modo che potremmo ordinare le narrazioni eterofenomenologiche in base al loro realismo, scartando quelle che asseriscono o presuppongono dei talenti discriminativi, o delle disposizioni reattive, di cui si può dimostrare l'assenza nella ecologia e nella neurofisiologia del pipistrello. Per esempio, impareremmo che i pipistrelli non sono infastiditi dai sonori squittii che emettono per produrre i loro echi, poiché dispongono di un muscolo sapientemente progettato che chiude le loro orecchie in perfetta sincronia con i loro squittii, in modo non dissimile ai congegni sincronizzati che permettono ai sensibili sistemi radar di evitare di essere disturbati dai propri segnali in uscita. Molte ricerche sono già state fatte su questi argomenti, così possiamo già dire molto di più, per esempio possiamo già dire perché i pipistrelli usano differenti configurazioni di frequenze nei loro squittii, a seconda se stanno scandagliando l'ambiente alla ricerca di una preda, o stanno avvicinandosi ad un obiettivo o apprestandosi ad uccidere (Akins, 1989, 1990).

Quando arriviamo a narrazioni eterofenomenologiche in cui nessun critico può trovare aspetti fondatamente criticabili, noi dovremmo accettarle – provvisoriamente, in attesa di altre scoperte – come spiegazioni accurate di che cosa si prova ad essere la creatura in questione. Questo, dopo tutto, è il modo in cui ci trattiamo l'un l'altro. Raccomandando di trattare nello stesso modo anche i pipistrelli e altri candidati all'interpretazione, io non sto *sottraendomi* all'onere della prova, sto solo

estendendo ad altre entità il normale, umano, onere della prova.

Potremmo usare queste indagini per far svanire ogni sorta di illusione apertamente romantica sulla coscienza dei pipistrelli. Noi *sappiamo* che il delizioso libro per bambini di Randall Jarrell, *The Bat-Poet* (1963), è solo frutto della fantasia, perché sappiamo che i pipistrelli non parlano! Affermazioni meno ovviamente fantastiche sulla loro fenomenologia soccombono di fronte a fatti meno ovvi, ma pur sempre pubblici, sulla loro fisiologia e il loro comportamento. Queste indagini ci darebbero molte informazioni su ciò di cui un pipistrello potrebbe o non potrebbe essere cosciente in varie condizioni, mostrandoci quali mezzi ci siano nei loro sistemi nervosi per rappresentare questo e quello, e controllando sperimentalmente che il pipistrello usi realmente quell'informazione nella modulazione del suo comportamento. È difficile immaginare quanto si possa racimolare da questo tipo di ricerca fino a che non si prova. (Per un'indagine preliminare sorprendentemente dettagliata su cosa si prova ad essere un cercopiteco, per esempio, vedi Cheney e Seyfarth, *How Monkeys See the World*, 1990.)

Tutto ciò suggerisce un'ovvia obiezione: queste indagini ci darebbero molte notizie sull'organizzazione cerebrale e l'elaborazione dell'informazione nel pipistrello, ma ci direbbero solamente di cosa i pipistrelli *non* sono coscienti, lasciando del tutto irrisolta la questione di cosa i pipistrelli *siano* coscienti (se sono coscienti). Come sappiamo, gran parte dell'elaborazione dell'informazione nei sistemi nervosi è completamente inconscia, quindi questi metodi di indagine non sarebbero in grado di escludere l'ipotesi che i pipistrelli siano... zombi volanti, e che quindi non si provi nulla ad essere un pipistrello! (Wilkes, 1988, p. 224, si chiede se l'ecolocazione non sia una specie di visione cieca, che quindi non si prova nulla ad avere.)

Aha, il mistero è svelato. Ecco la malaugurata direzione in cui sta scivolando la discussione; ma noi dovremo impedirlo. Richard Dawkins (1986), in una discussione illuminante sul funzionamento dell'ecolocazione nei pipistrelli *rhinolophus ferrumequinum*, ci dà una versione chiara dell'immagine sottintesa.

La polizia sfrutta l'effetto Doppler per controllare la velocità degli automobilisti... Comparando la frequenza dei segnali in uscita con quella dell'eco di ritorno, i poliziotti, *o meglio i loro strumenti automatici* [il corsivo è mio], possono calcolare la velocità di ogni automobile... Comparando l'altezza delle sue grida con quella dell'eco di ritorno,

quindi, il pipistrello (*o meglio il suo calcolatore portatile nel cervello*) [il corsivo è mio] potrebbe, in teoria, calcolare a che velocità si muove verso l'albero. (pp. 30-31)

Verrebbe da chiedersi: c'è qualcosa nel pipistrello che occupa una posizione rispetto al *suo* «calcolatore portatile» (che opera senza l'ombra di coscienza) identica a quella della polizia rispetto ai loro «congegni automatici»? I poliziotti non devono calcolare consciamente lo spostamento Doppler, ma devono avere l'esperienza, cosciente, del display del loro congegno che dice, in simboli luminosi a LED rossi: «150 km orari». Questo è ciò che li fa saltare sulle loro motociclette e inserire le sirene. È plausibile supporre che neanche il pipistrello debba calcolare consciamente lo spostamento Doppler – il suo calcolatore portatile si occupa di ciò – ma allora non si apre la possibilità che nel pipistrello ci sia qualcosa di analogo ai poliziotti che hanno l'esperienza, una sorta di testimone che apprezzi (consciamente) «l'output» del calcolatore del pipistrello che analizza l'effetto Doppler? Si tenga presente che potremmo rimpiazzare abbastanza facilmente i poliziotti con un congegno automatico che in qualche maniera registrasse la targa del veicolo trasgressore, cercasse il nome e l'indirizzo del proprietario e gli inviasse la multa. Non c'è nulla di speciale nel compito portato a termine dai poliziotti che impedisca di pensare che possa essere eseguito senza la minima coscienza. Lo stesso sembrerebbe valido anche per i pipistrelli. Un pipistrello potrebbe essere uno zombi. Sarebbe uno zombi – suggerisce questa linea di ragionamento – a meno che non ci fosse un osservatore al suo interno che reagisse ad una presentazione interna in modo molto simile a quello in cui i poliziotti reagiscono alla luce rossa lampeggiante dei loro strumenti.

Non cadere nella trappola. Questa è la nostra vecchia nemesi, il Pubblico nel Teatro Cartesiano. La *tua* coscienza non consiste nel fatto che il tuo cervello sia abitato da un agente interiore a cui il tuo cervello presenta uno spettacolo, così la nostra incapacità di trovare un tale agente centrale nel cervello del pipistrello non metterebbe in dubbio la tesi della sua coscienza, né la tesi della nostra capacità a dire com'è la sua coscienza. Per comprendere la coscienza del pipistrello, dobbiamo semplicemente applicare al pipistrello gli stessi principi che applichiamo a noi stessi.

Ma cosa potrebbe fare, allora, un pipistrello per convincerci che siamo in presenza di un'autentica coscienza? Può sembrare

che anche mettendo qualsiasi utilizzatore sofisticato dell'output dietro i trasduttori Doppler del pipistrello, non avremmo alcuna ragione convincente – esterna, «in terza persona» – per garantire al pipistrello un'esperienza cosciente. Non è così. Se il pipistrello potesse parlare, per esempio, potrebbe produrre un testo dal quale potremmo ricavare un mondo eterofenomenologico, e questo ci darebbe le stesse identiche basi per garantirgli la coscienza di quelle che servono per qualunque altra persona. Ma, come abbiamo già notato, i pipistrelli non parlano. Possono comportarsi, comunque, in molte maniere non verbali che possono fornire una base chiara per descrivere il loro mondo eterofenomenologico o, come il pionieristico ricercatore von Uexküll (1909) lo ha chiamato, il loro *Umwelt und Innenwelt*, il loro Mondo Circostante e Mondo Interiore.

L'eterofenomenologia senza un testo non è impossibile, solo difficile (Dennett, 1988a, 1988b, 1989a, 1989b). Un ramo dell'eterofenomenologia animale è conosciuto come etologia cognitiva, il tentativo di costruire dei modelli delle menti degli animali studiando – e sperimentando su – il loro comportamento sul campo. Le possibilità e le difficoltà di questo tipo di indagine sono evidenziate bene in Cheney e Seyfarth (1990), Whiten e Byrne (1988) e in Ristau (1991), una raccolta di scritti in onore di Donald Griffin, il pioniere delle ricerche sulla ecolocazione dei pipistrelli e il creatore del campo dell'etologia cognitiva. Una delle difficoltà più frustranti incontrate da questi ricercatori è la completa impraticabilità di molti esperimenti desiderati a causa dell'assenza di linguaggio; non è semplicemente possibile *preparare* i soggetti (e sapere che si è riusciti a prepararli) nei modi richiesti da questi esperimenti senza conversare con i soggetti (Dennett, 1988a).

Questo non è solo un problema epistemologico per l'eterofenomenologo; la stessa difficoltà di creare nell'ambiente naturale le circostanze sperimentali richieste dimostra qualcosa di più fondamentale sulle menti delle creature prive di linguaggio. Mostra che le situazioni ecologiche di questi animali non hanno mai fornito loro le *opportunità* di sviluppare (tramite l'evoluzione, tramite l'apprendimento o entrambi) molte delle attività mentali avanzate che forgiano le nostre menti, e quindi possiamo essere abbastanza sicuri che non le hanno mai sviluppate. Per esempio, considera il concetto di un *segreto*. Un segreto non è solo qualcosa che tu sai e gli altri no. Affinché tu abbia un segreto è necessario che tu sappia che gli altri non lo conoscono e tu devi essere in grado di controllare questo fatto. (Se tu sei il

primo a vedere una mandria che si avvicina al galoppo, puoi sapere qualcosa che gli altri non sanno, ma non per molto tempo; non puoi *tenere* segreta questa informazione privilegiata.) L'ecologia comportamentale di una specie deve essere strutturata in modo piuttosto speciale affinché ci sia un qualche ruolo per i segreti. Le antilopi, nelle loro mandrie, non hanno segreti e non hanno il modo di averli. Così un'antilope probabilmente non è capace di tramare un piano segreto più di quanto non lo sia di contare fino a cento o di godere dei colori del tramonto. I pipistrelli, che praticano solitarie scorrerie durante le quali potrebbero essere in grado di riconoscere tale isolamento dai loro rivali, soddisfano una delle condizioni necessarie per avere segreti. Essi hanno anche interessi che potrebbero essere chiaramente ben serviti sfruttando i segreti? (Cosa ci farebbe una vongola con un segreto? Se ne starebbe semplicemente a ridere tra sé nella sabbia?) I pipistrelli usano nella loro caccia delle abitudini di segretezza e inganno che potrebbero essere incorporate in attività più elaborate volte a tenere segreti? Ci sono effettivamente molte questioni di questo tipo che, una volta sollevate, suggeriscono ulteriori ricerche ed esperimenti. La struttura della mente di un pipistrello è altrettanto accessibile della struttura del suo sistema digestivo; il modo per indagarli entrambi è quello di andare sistematicamente avanti e indietro tra un'analisi dei suoi contenuti e un'analisi del mondo dal quale il suo contenuto era derivato, prestando attenzione ai metodi e agli scopi della derivazione.

Wittgenstein una volta disse: «Se un leone potesse parlare, noi non potremmo capirlo» (1953, p. 292). Io penso, al contrario, che se un leone potesse parlare, quel leone avrebbe una mente così differente dall'andamento generale della mente dei leoni che, sebbene potremmo capirlo molto bene, da lui impareremmo ben poco sulla mente degli altri leoni. Il linguaggio, come abbiamo visto nei capitoli precedenti, gioca un ruolo enorme nella strutturazione di una mente umana; quindi non dovremmo supporre che la mente di una creatura priva di linguaggio – e che non ha in realtà alcun bisogno di esso – sia strutturata nelle stesse maniere. Questo significa che gli animali privi di linguaggio «non sono affatto coscienti»? (come pretendeva Cartesio). Questa domanda fa sorgere sempre a questo momento una sfida nei confronti delle nostre credenze, ma non dovremmo sentirci obbligati a darne una risposta così come è formulata. Si noti che essa presuppone qualcosa che abbiamo faticosamente evitato: la presupposizione che la co-

scienza sia una proprietà speciale «tutto o niente» che divide l'universo in due categorie profondamente differenti: le entità che la possiedono (le entità che si prova qualcosa ad essere, come direbbe Nagel) e le entità che non la possiedono. Persino nel nostro proprio caso, non possiamo tracciare la linea che separa i nostri stati mentali coscienti dai nostri stati mentali non coscienti. La teoria della coscienza che abbiamo delineato permette un'ampia varietà nell'architettura funzionale e sebbene la presenza del linguaggio segni un incremento particolarmente significativo della gamma, della versatilità e dell'autocontrollo immaginativo (per menzionare solo alcuni dei poteri più ovvi della macchina virtuale *joyceana*), questi poteri non hanno l'*ulteriore* potere di far accendere una particolare luce interiore che sarebbe altrimenti spenta.

Quando immaginiamo che cosa si prova ad essere una creatura priva di linguaggio, noi iniziamo, naturalmente, dalla nostra propria esperienza e la maggior parte di quello che ci viene allora in mente deve essere riaggiustato (soprattutto ridimensionato). Il tipo di coscienza di cui godono tali animali è drammaticamente tronca, se paragonata alla nostra. Un pipistrello, per esempio, non solo non può chiedersi se è venerdì; non può neanche chiedersi se è un pipistrello; non c'è un ruolo per l'interrogarsi nella sua struttura cognitiva. Sebbene un pipistrello, come anche l'umile aragosta, abbia un sé biologico, non ha un vero sé di cui parlare – nessun Centro di Gravità Narrativa, o al massimo uno molto trascurabile. Nessuna «parola sulla punta della lingua», ma anche nessun rimpianto, nessuna complessa bramosia, nessuna reminiscenza nostalgica, nessuno schema profondo, nessuna riflessione su cosa si prova ad essere un gatto o perfino su cosa si prova ad essere un pipistrello. Questo elenco di negazioni equivarrebbe ad uno scetticismo a buon mercato se non avessimo una teoria empirica precisa su cui basarci. Sto sostenendo di aver dimostrato che i pipistrelli *non possono* avere questi stati mentali? Beh, no, ma non posso neanche provare che i funghi *non siano* astronavi intergalattiche inviate a spiarci.

Questo non è un orrendo pregiudizio antropocentrico? E che dire, poi, dei sordomuti? Non sono coscienti? È naturale che sono coscienti – ma evitiamo di saltare a stravaganti conclusioni sulla loro coscienza, basate su una fuorviante simpatia. Quando un sordomuto acquista un linguaggio (in particolare, il linguaggio dei segni, il linguaggio più naturale che un sordomuto possa imparare), nasce una mente umana completamente sviluppata,

chiaramente e dimostrabilmente differente dalla mente di una persona senza problemi di udito, ma ugualmente capace di tutte le intricatezze riflessive e i poteri generativi – e forse anche di più. Ma senza un linguaggio naturale, la mente di un sordomuto è terribilmente handicappata. (Vedi Sacks, 1989, specialmente la bibliografia ragionata.) Come il filosofo Ian Hacking (1990) nota nella recensione al libro di Sacks, «Serve una fervida immaginazione perfino per avere una vaga idea di quello che sta perdendo un bambino sordo». Non si fa un favore ai sordomuti immaginando che in assenza di linguaggio godano di tutte le delizie mentali di cui godiamo noi esseri umani udenti e non si fa un favore agli animali non umani cercando di nascondere i fatti disponibili sulle limitazioni delle loro menti.

E questo, come molti si affretteranno a precisare, è qualcosa che si poteva cogliere fra le righe da parecchio tempo e solo ora è emerso chiaramente: molta gente è spaventata dalla prospettiva che la coscienza venga spiegata perché temono che, se ci si riuscisse, perderemmo le nostre coordinate morali. Forse *possiamo* immaginare un calcolatore cosciente (o la coscienza di un pipistrello) ma *non dovremmo provarci*. Se prendiamo questa abitudine negativa, cominceremo a trattare gli animali come se fossero giocattoli a molla, infanti e sordomuti come se fossero orsacchiotti di peluche e – per aggiungere la beffa al danno – robot come se fossero persone reali.

3. *Preoccuparsi e dare importanza*

Il titolo di questo paragrafo deriva da un articolo di Marian Stamp Dawkins (1987), che ha svolto delle ricerche accurate sulle implicazioni morali dell'eterofenomenologia animale. (Il suo lavoro precedente è contenuto nel libro *Animal Suffering: The Science of Animal Welfare*, 1980.) Come la Dawkins nota, i nostri atteggiamenti morali verso gli animali sono pieni di incoerenze.

> Per evidenziare le nostre incoerenze non dobbiamo far altro che pensare a vari tipi di animali. Ci sono innumerevoli manifestazioni contro l'uccisione delle foche neonate, ma nessuna campagna del genere per fermare l'uccisione dei ratti. Molta gente non ha alcun problema a mangiare carne di maiale o di pecora, ma è inorridita all'idea di mangiare carne di cane o di cavallo. [p. 150]

Dawkins indica che ci sono due fili principali in questo

groviglio: la capacità di ragionare e la capacità di soffrire. Cartesio ha dato molta importanza all'incapacità degli animali non umani di ragionare (almeno nel modo in cui ragionano gli esseri umani), e ciò ha provocato una famosa risposta da parte del filosofo utilitarista britannico Jeremy Bentham: «un cavallo, o un cane, adulto è un animale incomparabilmente più razionale, ed anche più comunicativo, di un infante di un giorno o una settimana o anche un mese. Ma anche se fosse il contrario, a che servirebbe? L'interrogativo non è "possono *ragionare*?" o "possono *parlare*?" ma "possono *soffrire*?"» (Bentham, 1789). Di solito questi sembrano criteri opposti di status morale, ma come Dawkins sostiene, «dare un valore etico alla capacità di soffrire ci porterà alla fine a dare valore agli animali intelligenti. Anche se partiamo rifiutando il criterio del ragionamento di Cartesio, sono gli animali in grado di ragionare quelli che più verosimilmente possiedono la capacità di soffrire» (p. 153).

Le ragioni per giustificare questa affermazione sono implicite nella teoria della coscienza che ho sviluppato. Soffrire non significa essere visitato da qualche stato ineffabile ma intrinsecamente orrendo, significa invece avere le proprie speranze vitali, i propri piani vitali, i propri progetti vitali frustrati dalle circostanze che si sovrappongono ai propri desideri, vanificando le proprie intenzioni – quali che siano. L'idea che la sofferenza sia in un certo senso spiegabile dalla presenza di qualche proprietà intrinseca – diciamo, la «terribilezza» – è senza speranza così come è senza speranza l'idea che il divertimento sia in un certo senso spiegabile dalla presenza di una comicità intrinseca. Così la presunta inaccessibilità, la inconoscibilità ultima, della sofferenza di un altro è altrettanto fuorviante delle altre fantasie sui qualia intrinseci che abbiamo già smascherato, sebbene meno ovviamente perniciose. Ne consegue – e ciò tocca una corda intuitiva – che la capacità di soffrire è una funzione della capacità di avere desideri, aspettative ed altri sofisticati stati mentali che siano altamente discriminativi, articolati e variegati.

Gli esseri umani non sono le uniche creature abbastanza intelligenti per soffrire; il cavallo e il cane di Bentham mostrano attraverso il loro comportamento di avere una complessità mentale sufficiente per distinguere – e preoccuparsi di – uno spettro di dolori e altri abusi tutt'altro che trascurabile, anche se appare una ben angusta finestra se paragonato al campo delle possibilità della sofferenza umana. Altri mammi-

feri, principalmente le scimmie, gli elefanti e i delfini, sembrano avere delle gamme molto più ampie.

A compenso di tutta la sofferenza che devono sopportare, le creature intelligenti ottengono tutto il divertimento. Devi avere un'economia cognitiva che preveda l'esplorazione e l'autostimolazione per disporre dello spazio necessario all'accumulazione ricorsiva dei desideri derivati che rende possibile il divertimento. Il primo passo avviene quando la tua architettura ti permette di apprezzare il significato di «Così, così, mi piace tanto!». Deboli versioni di questo potere costruttivo sono manifeste in alcune specie superiori, ma serve un'immaginazione rigogliosa e del tempo libero – cosa che molte specie non possono permettersi – per sviluppare un ampio spettro di piaceri. Più grande è il campo, più ricchi i dettagli, più sottilmente discriminativi i desideri, e peggio si sta quando quei desideri vengono frustrati.

Ma che importanza dovrebbe avere, potresti chiedere, che i desideri di una creatura siano frustrati se non sono desideri *coscienti*? La mia replica: perché dovrebbe importare di più nel caso di desideri coscienti – specialmente se la coscienza viene considerata una proprietà, come qualcuno pensa, che sfugge sempre all'indagine? Perché le speranze infrante di uno «zombi» dovrebbero essere meno importanti delle speranze infrante di una persona cosciente? Qui c'è un trucco ben fatto che dovrebbe essere svelato e rifiutato. La coscienza, dici, è ciò che conta, ma poi ti aggrappi a dottrine sulla coscienza che ci impediscono sistematicamente di ottenere qualsiasi ragguaglio sul *perché* essa conta. Postulare speciali qualità interne che non sono solo private e intrinsecamente preziose, ma anche non confermabili e non indagabili è solo oscurantismo.

Dawkins mostra come le differenze indagabili – le uniche differenze che possano importare – possono essere esplorate sperimentalmente, e vale la pena soffermarsi un po' sui dettagli per far vedere quante conoscenze possiamo racimolare anche da semplici esperimenti con specie non troppo attraenti.

Le galline lasciate libere o tenute in grandi pollai con lettiere passano la maggior parte del tempo vagando e razzolando, quindi ho creduto che la mancanza di lettiera nelle gabbie in batteria potesse causare una sofferenza alle galline. Sicuramente, quando ho dato loro la possibilità di scegliere tra una gabbia con il fondo in rete metallica e una con il fondo ricoperto da lettiera su cui potevano razzolare, esse sceglievano la gabbia con lettiera. Infatti entravano anche in una

gabbia molto angusta (così piccola che si potevano a mala pena girare), se questo era l'unico modo per accedere alla lettiera. Anche dei volatili che sono stati allevati per tutta la loro vita in gabbie e non hanno mai avuto prima l'esperienza della lettiera scelgono la gabbia con la lettiera sul fondo. Benché ciò sembrasse già un forte indizio, non era ancora abbastanza. Dovevo mostrare non solo che le galline avevano una spiccata preferenza per la lettiera ma anche che avevano una preferenza abbastanza forte da poter dire che forse soffrono se ne vengono private.

Alle galline è stata offerta, quindi, una scelta leggermente differente. Questa volta dovevano scegliere tra una gabbia con il fondo in rete metallica in cui erano posti acqua e cibo e una gabbia con lettiera senza acqua e cibo... Il risultato fu che esse passavano la maggior parte del tempo nella gabbia con lettiera e molto meno tempo nella gabbia con la rete metallica, anche se questo era l'unico posto dove potevano bere e mangiare. Poi fu introdotta una complicazione. I volatili dovevano «lavorare» per muoversi tra le gabbie. Dovevano o saltare da un corridoio o passare attraverso un divisorio di plastica nera. Così il passaggio da una gabbia all'altra ora aveva un costo... Le galline seguitavano a passare la stessa quantità di tempo nella gabbia in rete metallica con il cibo, come prima quando non c'era difficoltà ad entrarci. Ma non stavano quasi mai nella gabbia con lettiera. Non sembravano, semplicemente, preparate a lavorare o pagare un costo per entrare nella gabbia con lettiera... Al contrario di quello che mi aspettavo, i volatili sembravano dire che la lettiera non era davvero importante per loro. [pp. 157-159]

Dawkins conclude dicendo che «La sofferenza della mente emozionale è rivelata da animali che hanno una mente razionale sufficientemente sviluppata da essere in grado di agire sulle condizioni che li fanno soffrire», e prosegue notando che «è anche verosimile che organismi senza la capacità di agire per sottrarsi alla fonte del pericolo non svilupperanno la capacità di soffrire. Non ci sarebbe nessuna ragione evoluzionistica che un albero a cui stanno segando i rami abbia la capacità di soffrire in silenzio» (p. 159). Come abbiamo visto nel capitolo 7 (vedi anche capitolo 3, nota 6), si deve fare molta attenzione nel formulare questi argomenti evoluzionistici sulla funzione, perché la storia gioca un grande ruolo nell'evoluzione, e la storia spesso gioca brutti scherzi. Ma in assenza di basi precise per attribuire la sofferenza, o in assenza di basi precise per sospettare che le suddette basi precise siano per una ragione o per l'altra sistematicamente oscurate, dovremmo concludere che non c'è sofferenza. Non dobbiamo aver paura che questa regola austera ci conduca a trascurare i nostri obblighi nei confronti

delle altre creature. Essa ci fornisce anche una base ampia per delle conclusioni precise: molti, ma non tutti, gli animali sono capaci di gradi significativi di sofferenza. Un'argomentazione più convincente a favore di un trattamento umanitario può essere costruita riconoscendo le grandi differenze di grado, piuttosto che promulgando un ingiustificabile dogma sull'universalità e sull'uguaglianza del dolore animale.

Ciò può risolvere la questione oggettiva sulla presenza o l'assenza della sofferenza, ma non calma i sentimenti morali turbati dalla prospettiva di una spiegazione della coscienza in modi così insensibilmente meccanici. C'è qualche altra cosa in gioco.

Io ho una fattoria nel Maine e sono contento che ci siano orsi e coyote che vivono nei miei boschi. Capita molto raramente che io ne veda qualcuno – o anche solo i segni della loro presenza – ma mi piace sapere che sono lì e mi dispiacerebbe molto venire a sapere che se ne sono andati. E non mi sentirei completamente compensato della perdita, se qualcuno dei miei amici del campo dell'IA riempisse i miei boschi con innumerevoli bestie robotiche (sebbene, a rifletterci bene, l'idea sia affascinante). È importante per me che ci siano creature selvagge, discendenti di creature selvagge, che vivono così vicino a me. Similmente, mi fa piacere che ci siano concerti che si tengono nell'area di Boston che io non ascolto o di cui non sento neanche parlare.

Questi fatti sono di un tipo speciale. Sono fatti che sono importanti per noi semplicemente perché una parte dell'ambiente che conta per noi è l'ambiente delle nostre credenze. E poiché è difficile che noi continuiamo a credere a delle affermazioni la cui giustificazione è svanita, è importante per noi che le credenze siano *vere*, anche quando noi non vedremo direttamente nessuna prova a loro sostegno. Come ogni altra parte dell'ambiente, un ambiente di credenze può essere fragile, composto di parti che risultano interconnesse tramite sia accidenti storici sia legami ben progettati. Si consideri, per esempio, quella parte delicata del nostro ambiente di credenze che riguarda le disposizioni sui nostri corpi dopo la morte. Poche persone credono che l'anima rimane nel corpo dopo la morte – anche le persone che credono nell'anima non credono *ciò*. Eppure ben pochi di noi tollererebbero una «riforma» che incoraggiasse le persone a sbarazzarsi dei loro parenti morti mettendoli nei sacchi di plastica della spazzatura o in un modo altrettanto poco cerimonioso. Perché no? Non perché noi cre-

diamo che i cadaveri possano effettivamente soffrire di qualche offesa. Un cadavere non può patire un'offesa più di quanto non lo possa un tronco d'albero. Eppure, l'idea è scioccante, repellente. Perché?

Le ragioni sono complesse, ma possiamo distillare alcuni punti che ci interessano ora. Una persona non è solo un corpo; una persona *ha* un corpo. Quel cadavere è il corpo del caro vecchio Bianchi, un Centro di Gravità Narrativa che deve la sua realtà sia al corpo che giace ora senza vita sia agli sforzi con cui noi collaboriamo ad una reciproca interpretazione eterofenomenologica. I confini di Bianchi non sono identici ai confini del corpo di Bianchi, e gli interessi di Bianchi, grazie alla curiosa pratica umana di auto-tessitura, possono estendersi oltre i basilari interessi biologici che generarono tale pratica. Noi trattiamo il suo cadavere con rispetto perché è importante per la conservazione dell'ambiente di credenze in cui tutti viviamo. Se cominciamo a trattare i cadaveri come spazzatura, per esempio, potrebbe cambiare il modo con cui trattiamo i quasi-cadaveri – quelli che sono ancora vivi ma stanno per morire. Se non errassimo sul versante del prolungamento dei rituali e delle pratiche di rispetto ben oltre la soglia della morte, i morenti (e i loro cari) si troverebbero di fronte ad un'ansietà, un affronto, una *possibilità*, che rischia di offenderli. Trattare «malamente» un corpo potrebbe non far male a nessuna persona morente, e certamente non fa male a nessun cadavere, ma, se diventasse una pratica comune e fosse ampiamente risaputa (come accadrebbe sicuramente), ciò cambierebbe significativamente l'ambiente di credenze che circonda il morire. La gente immaginerebbe gli eventi successivi alla loro dipartita in un modo differente da come li immagina ora, e sarebbe un modo particolarmente deprimente. Forse ingiustificatamente, ma cosa importa? Se la gente rischia di diventare depressa, questa è di per sé una buona ragione per non adottare una pratica.

Così ci sono ragioni indirette, ma ancora credibili, legittime e robuste per continuare a rispettare i cadaveri. Non abbiamo bisogno di alcuna mitologia su qualcosa di speciale che risieda effettivamente nei cadaveri e che li renda privilegiati. Questo *potrebbe* essere un mito utile da diffondere tra i più semplici, ma sarebbe estremamente paternalistico pensare che i più informati tra noi debbano conservare tali miti. Analogamente, esistono delle ragioni perfettamente buone per trattare tutti gli animali viventi con cura e sollecitudine. Queste ragioni sono in un certo senso indipendenti dai fatti relativi alla questione di

che tipo di dolore senta ogni tipo di animale. Esse dipendono più direttamente dal fatto che varie credenze sono diffuse nella nostra cultura, e hanno importanza per noi, a prescindere da quanto *dovrebbero* avere importanza o no. Poiché sono importanti ora, esse sono importanti. Ma la razionalità dell'ambiente di credenze – il fatto che le credenze stupide o senza base tendono ad estinguersi alla lunga, malgrado la superstizione – implica che le cose importanti ora possono non essere sempre importanti.

Ma allora, come abbiamo anticipato nel capitolo 2, una teoria che attacca radicalmente l'ambiente generale di credenze è dotata di un potenziale autentico di dolore, di sofferenza (per la gente che si preoccupa particolarmente degli animali, per esempio, indipendentemente dal fatto se ciò che accade agli animali equivalga alla sofferenza). Ciò significa che dobbiamo smetter di indagare questi argomenti, per paura di scoperchiare il vaso di Pandora? Questo potrebbe essere giustificato, se potessimo convincerci che il nostro attuale ambiente di credenze, dominato dai miti o no, sia un ambiente benigno e moralmente accettabile, ma io credo che sia chiaro che non è così. Coloro che si preoccupano dei costi minacciati da questo illuminismo non richiesto dovrebbero guardare con più attenzione ai costi dei miti correnti. Pensiamo realmente che ciò che oggi ci troviamo di fronte sia degno di essere protetto da qualche oscurantismo creativo? Pensiamo, per esempio, che dovremmo destinare vaste risorse per conservare le prospettive immaginarie di una vita mentale rinnovata per le persone in coma profondo, mentre non ci sono risorse disponibili per migliorare le aspettative disperate, ma tutt'altro che immaginarie, dei poveri? Miti sulla santità della vita, o della coscienza, sono armi a doppio taglio. Possono essere utili per erigere barriere (contro l'eutanasia, contro la pena capitale, contro l'aborto, contro il mangiare carne), per impressionare i sempliciotti, ma al prezzo di un'offensiva ipocrisia o di un ridicolo auto-inganno tra i più illuminati.

Le barriere assolute, come la Linea Maginot, raramente servono allo scopo per cui erano state progettate. Le campagne che venivano una volta intraprese contro il materialismo hanno già ceduto alle difficoltà e le campagne contro «l'IA forte», sebbene ugualmente ben motivate, possono offrire solo i più triti modelli alternativi della mente. Sicuramente sarebbe meglio tentare di incoraggiare un atteggiamento che fondi la sollecitudine morale su basi *non* assolutistiche, *non* intrinseche, *non*

dicotomizzate e che possa coesistere con la nostra crescente conoscenza del funzionamento interno di questa incredibile macchina, il cervello. Gli argomenti morali a favore di *entrambe* le posizioni sulla pena capitale, l'aborto, il mangiare carne e la vivisezione, per esempio, vengono portati ad un livello più alto, più appropriato quando scartiamo i miti che sono in ogni caso indifendibili.

4. *Coscienza spiegata o coscienza eliminata?*

Quando impariamo che l'unica differenza tra l'oro e l'argento consiste nel numero di particelle subatomiche nei loro atomi, ci possiamo sentire ingannati e arrabbiati – quei fisici hanno eliminato qualcosa: hanno fatto svanire la «aurità» dell'oro, hanno lasciato fuori proprio quella «argenteità» dell'argento che noi tanto apprezziamo. E quando spiegano il modo in cui la riflessione e l'assorbimento delle radiazioni elettromagnetiche rende ragione dei colori e della visione cromatica, sembrano dimenticare proprio la cosa che conta di più. Ma naturalmente ci deve essere qualcosa che viene «lasciato fuori» – altrimenti non avremmo iniziato a spiegare. Il fatto che qualcosa venga lasciato fuori non è una caratteristica delle spiegazioni mancate, ma delle spiegazioni riuscite.

Solo una teoria che spiega la coscienza in termini di eventi inconsci può spiegare la coscienza. Se il tuo modello della produzione del dolore da parte delle attività cerebrali presenta un riquadro con l'etichetta «dolore», non hai ancora iniziato a spiegare che cosa sia il dolore, e se il tuo modello della coscienza va avanti perfettamente fino al momento magico in cui devi dire «e poi avviene un miracolo», non hai iniziato a spiegare che cosa sia la coscienza.

Ciò conduce alcune persone ad asserire che la coscienza non può mai essere spiegata. Ma perché la coscienza dovrebbe essere l'unica cosa che non può mai essere spiegata? I solidi e i liquidi e i gas possono essere spiegati in termini di cose che non sono esse stesse solide o liquide o gassose. Sicuramente la vita può essere spiegata in termini di cose che non sono esse stesse vive – e la spiegazione non lascia senza vita le cose viventi. L'illusione che la coscienza costituisca l'eccezione deriva – io sospetto – da una incomprensione di questa caratteristica generale delle spiegazioni riuscite. Pensando, erroneamente, che la spiegazione lasci fuori qualcosa, pensiamo di salvare ciò che

andrebbe altrimenti perso rimettendolo nell'osservatore come un qualia – o come qualche altra proprietà «intrinsecamente» meravigliosa. La psiche diventa la sottana protettiva sotto cui si possono nascondere tutti questi amati gattini. Ci potrebbero essere dei *motivi* per credere che la coscienza non possa essere spiegata, ma io spero di aver mostrato che ci sono buone *ragioni* per pensare che possa essere spiegata.

La mia spiegazione della coscienza è tutt'altro che completa. Si potrebbe perfino dire che è stato solo un inizio, ma *è* un inizio, perché rompe l'incantesimo creato dalle idee che fanno sembrare impossibile una spiegazione della coscienza. Io non ho sostituito una teoria metaforica, il Teatro Cartesiano, con una teoria *non* metaforica («letterale, scientifica»). Tutto quello che ho fatto, realmente, è stato di sostituire una famiglia di immagini e metafore con un'altra: ho rimpiazzato il Teatro, il Testimone, l'Autore Centrale, il Figmento con il Software, le Macchine Virtuali, le Versioni Molteplici, un Pandemonio di Homunculi. È solo una guerra di metafore, potresti dire – ma le metafore non sono «solo» metafore; le metafore sono gli strumenti del pensiero. Nessuno può riflettere sulla coscienza senza di esse, così è importante equipaggiarsi con il migliore insieme disponibile di strumenti. Guarda che cosa abbiamo costruito con i nostri strumenti. Avresti mai potuto immaginarlo senza di essi?

APPENDICE A
(PER FILOSOFI)

Ci sono punti nel libro in cui io scavalco con un agile salto importanti battaglie filosofiche senza fare alcun commento o cercare in qualche modo di soddisfare le usuali obbligazioni di un filosofo accademico. I filosofi che hanno letto il manoscritto mi hanno fatto delle domande su queste lacune. Malgrado si riferiscano ad argomenti che forse non interessano i non filosofi, meritano una risposta.

Sembra che tu esegua un rapido gioco di prestigio alla fine del capitolo 11, nel dialogo con Otto, quando introduci brevemente i «presentimenti» come atti linguistici senza Attore e senza Discorso, e poi rivedi la tua auto-caricatura, sostituendo i presentimenti con «eventi di fissazione del contenuto» senza ulteriori spiegazioni. Non è questa la mossa cruciale di tutta la tua teoria?

Sì, è vero. Questo è il punto di contatto principale con l'altra metà della mia teoria della mente, la teoria del contenuto e dell'intenzionalità la cui presentazione più recente si trova in *The Intentional Stance.* Ci sono molti altri punti nel libro in cui mi baso su quella teoria, ma penso che tu abbia individuato il punto che sorregge il peso maggiore. Senza una teoria del contenuto, a quel punto la mia teoria avrebbe dovuto dire «E poi avviene un miracolo». La mia strategia fondamentale è sempre stata la stessa: prima, sviluppare una spiegazione del contenuto che sia *indipendente* dalla coscienza e *più basilare* di essa – una spiegazione del contenuto che si occupi in ugual maniera di tutte le fissazioni inconsce di contenuto (nei cervelli, nei calcolatori, nel «riconoscimento» da parte dell'evoluzione delle proprietà dei progetti selezionati) – e poi, costruire una spiegazione su quelle basi. Prima il contenuto, poi la coscienza. Le due metà di *Brainstorms* ricapitolano la strategia, ma le due metà della teoria sono cresciute e hanno fatto sorgere due volumi

differenti. Questo libro conclude la mia terza rassegna del problema. Naturalmente, questa strategia è completamente opposta alla visione di Nagel e di Searle, che in modi differenti insistono nel trattare la coscienza come basilare. La ragione per cui nel capitolo 11 ho scavalcato così rapidamente questo punto così centrale va ricercata semplicemente nel fatto che non riuscivo a vedere un modo in cui potessi comprimere le centinaia di pagine e di commenti dedicati alla teoria del contenuto in qualcosa che fosse soddisfacente sia per l'accuratezza che per l'accessibilità. Così se credi che in queste pagine io abbia eseguito un rapido gioco di prestigio, ti prego di consultare la versione più lenta nelle altre pagine citate in bibliografia.

Sembra esserci, comunque, una tensione – se non un'aperta contraddizione – tra le due metà della tua teoria. L'atteggiamento intenzionale presuppone (o incoraggia) la razionalità, e quindi l'unità, dell'agente – il sistema intenzionale – mentre il modello delle Molteplici Versioni si oppone in continuazione a questa unità centrale. Qual è, allora, la tua opinione sul modo giusto di concepire la mente?

Dipende tutto dalla distanza da cui ti poni. Più ti avvicini, più si rivela importante la disunità, la molteplicità, la competitività. Il mito del Teatro Cartesiano, dopo tutto, trae origine soprattutto da un'estrapolazione pigra ed *eccessiva* dell'atteggiamento intenzionale. Trattare un'entità complessa e mobile come un agente unitario è un'ottima maniera per rintracciare un ordine in tutta la sua attività; è una tattica che adottiamo naturalmente, e probabilmente è perfino favorita geneticamente come un modo di percepire e pensare. Ma se aspiriamo ad una scienza della mente, dobbiamo imparare a frenare e ridirigere queste abitudini di pensiero, frantumando l'agente unitario in mini-agenti e micro-agenti (senza un unico Boss). Allora possiamo capire come molti degli *apparenti* fenomeni dell'esperienza cosciente siano descritti erroneamente dalla tattica tradizionale, quella unitaria. Gli ammortizzatori della tensione sono le identificazioni più o meno forzate degli oggetti eterofenomenologici (così come sono concepiti nella prospettiva tradizionale) con gli eventi di fissazione del contenuto nel cervello (così come sono concepiti nella nuova prospettiva).

I filosofi hanno spesso evidenziato le idealizzazioni della tattica tradizionale, ma meno di frequente sono venuti a patti con esse. Per esempio, un'estesa letteratura filosofica è stata dedicata alle difficoltà della logica dei casi riflessivi della cre-

denza e della conoscenza, a cominciare da Hintikka (1962). Una delle idealizzazioni essenziali della formalizzazione di Hintikka, come egli ha esplicitamente chiarito, consisteva nella necessità che le affermazioni regolate dalla logica da lui presentata «siano fatte in *un'unica identica occasione*... La nozione del dimenticare non è applicabile entro i limiti di un'occasione» (p. 7). L'importanza di questa limitazione, egli notava, non è stata sempre apprezzata – e in genere si è persa nelle nebbie delle successive dispute. Hintikka riconosceva che questa quantificazione delle «occasioni» era una semplificazione necessaria per formalizzare i familiari concetti di credenza e conoscenza nel modo da lui fatto; il contenuto viene fissato in un istante e quindi viene fissata l'identità della proposizione in questione. Io ho sostenuto qui che questa individuazione artificiale in «stati» e «tempi» è una delle caratteristiche che trasformano in fantasie questi concetti della psicologia del senso comune quando proviamo a sovrapporli alle complessità di quello che avviene nel cervello.

In definitiva, cos'è per te un'esperienza cosciente? Sei un sostenitore della teoria dell'identità, sei un materialista eliminativo, un funzionalista, uno strumentalista?

Resisto alla richiesta di formulare una proposizione unica, formale e adeguatamente quantificata che riassuma in modo epigrammatico la mia teoria. Non ritengo che riempire la formula (x) (*x è un'esperienza cosciente se e solo se ...*) e difenderla dai controesempi proposti sia un buon metodo per sviluppare una teoria della coscienza, e credo di aver mostrato a sufficienza il perché. È proprio il modo indiretto del metodo eterofenomenologico che ci permette di scansare gli obblighi mal motivati di «identificare» o «ridurre» le (presunte) entità che abitano l'ontologia dei soggetti. Gli antropologi *identificano* Phaenomeno con il tizio che – hanno scoperto – faceva tutte le opere buone nella giungla, o sono «eliminativisti» rispetto a Phaenomeno? Se hanno fatto bene il *loro* lavoro, l'unico problema lasciato aperto è quello che può essere affrontato sulla base di considerazioni diplomatiche, non filosofiche o scientifiche. In un certo senso, si può dire che la mia teoria identifica le esperienze coscienti con gli eventi portatori di informazione nel cervello – poiché questo è tutto ciò che avviene e molti degli eventi cerebrali hanno una sorprendente somiglianza con gli abitatori dei mondi eterofenomenologici dei soggetti. Ma si potrebbero giu-

dicare «essenziali» altre proprietà degli oggetti eterofenomenologici – come la loro posizione nella sequenza temporale soggettiva, nel qual caso *non possono* essere identificati con gli eventi cerebrali disponibili, che possono essere in una sequenza differente, senza violare la Legge di Leibniz.

La questione se sia giusto trattare parte del mondo eterofenomenologico di un soggetto come un'utile finzione piuttosto che come una verità, più o meno forzata, non merita sempre molta attenzione. Le immagini mentali sono reali? Vediamo. Nel cervello della gente ci sono strutture reali di dati che sono piuttosto simili alle immagini – *esse* possono essere considerate le immagini di cui parli? Se sì, allora anche la risposta alla tua domanda è sì; se no, allora no. I qualia sono definibili funzionalmente? No, perché non ci sono proprietà come i qualia. O no, perché i qualia sono proprietà disposizionali dei cervelli che non sono strettamente definibili in termini *funzionali.* O sì, perché se comprendi realmente ogni cosa sul funzionamento del sistema nervoso, comprendi ogni cosa sulle proprietà di cui parla la gente quando pretende di parlare dei qualia.

Sono, allora, un funzionalista? Sì e no. Non sono un funzionalista delle macchine di Turing, ma allora dubito che ce ne siano mai stati; ed è un peccato, perché così vanno sprecate innumerevoli confutazioni. Io sono, naturalmente, una sorta di «teleofunzionalista», forse il primo teleofunzionalista (in *Contenuto e coscienza*), ma come ho sempre chiarito, ed enfatizzato qui nella discussione sull'evoluzione e sui qualia, non faccio l'errore di provare a definire tutte le differenze mentali rilevanti in termini di *funzioni* biologiche. Questa sarebbe una cattiva lettura di Darwin.

Sono uno strumentalista? Ritengo di aver mostrato perché questa sia una domanda mal concepita in «Real Patterns» (1991a). I dolori sono reali? Hanno la stessa realtà delle acconciature e dei dollari e delle opportunità e delle persone, e dei centri di gravità; ma quanta realtà è questa? Tutte queste domande dicotomizzanti sorgono dalla richiesta di riempire gli spazi della formula quantificata di prima, e alcuni filosofi pensano che una teoria della mente venga sviluppata mettendo assieme una proposizione blindata di quel tipo e poi difendendola. Una proposizione unica non è una teoria, è uno slogan; e ciò che alcuni filosofi fanno non è teorizzare, ma affinare slogan. A che serve questa fatica? Un successo in questa impresa quale confusione dissiperebbe, quali prospettive aprirebbe? Hai davvero bisogno di qualcosa da stampare sulla tua maglietta?

Alcuni affinatori di slogan sono davvero molto, molto bravi in questo, ma, come lo psicologo Donald Hebb una volta disse memorabilmente, «Se non vale la pena farlo, non vale la pena farlo bene».

Non intendo con ciò implicare che l'attenta definizione, e la critica delle definizioni tramite controesempi, non sia mai un utile esercizio. Si consideri, per esempio, la definizione del colore. Le ultime analisi e le definizioni recenti che i filosofi hanno tentato di dare ci hanno aperto gli occhi; hanno realmente chiarito i concetti ed evitato autentici equivoci. Vista l'attenzione che i filosofi hanno riservato recentemente al tentativo di dare una precisa definizione del colore, allora, la mia rapida affermazione, nel capitolo 12, che i colori sono «proprietà rifrangenti delle superfici degli oggetti o di volumi trasparenti» è estremamente semplicistica. Esattamente *quali* proprietà rifrangenti? Penso di aver spiegato perché sarebbe uno spreco di tempo cercare di dare una risposta precisa a tale domanda; l'unica risposta precisa non potrebbe essere una risposta concisa, per motivi che possiamo capire bene. Ciò significa che una definizione «non circolare» è difficile da ottenere. E allora? Penso davvero che questa semplice mossa sia in grado di fronteggiare i problemi sollevati dalla competizione? (In aggiunta a quelli citati in precedenza, vorrei menzionare Strawson, 1989, e Boghossian e Velleman, 1989, 1991.) Sì, ma è una storia lunga, così mi limiterò a rimettere la palla nel loro campo.

La tua posizione, in definitiva, non è solamente un tipo di verificazionismo?

I filosofi sono riusciti ultimamente a convincere se stessi – e molti spettatori innocenti – che il verificazionismo sia *sempre* un peccato. Sotto l'influenza di Searle e Putnam, per esempio, il neuroscienziato Gerald Edelman si ritrae precipitosamente da un atto di quasi-verificazionismo: «L'assenza di indizi di autocoscienza in animali differenti dagli scimpanzé non ci permette di considerarli non autocoscienti» (1989, p. 280) Vergogna! Un po' di coraggio! Sicuramente possiamo non solo *considerare* che non lo siano, ma possiamo approfondire questa idea, e se troviamo motivi validi e precisi per negare l'autocoscienza, dovremmo negarla. È tempo che il pendolo oscilli nell'altra direzione. Commentando una mia critica precedente a Nagel (Dennett, 1982a), Richard Rorty una volta disse:

Dennett pensa che si possa essere scettici sull'insistenza di Nagel ad attribuire una vita interiore fenomenologicamente ricca ai pipistrelli «senza per questo diventare il Verificazionista del Villaggio». Io no. Io penso che lo scetticismo su intuizioni come quelle di Nagel e di Searle è plausibile solo se è basato su considerazioni metodologiche generali sullo status delle intuizioni. Il verificazionista si lamenta in generale con il realista perché insiste su differenze (per esempio, tra pipistrelli con vita privata e pipistrelli senza vita privata, tra cani con intenzionalità intrinseca e cani senza intenzionalità intrinseca) che non fanno differenza: perché le sue intuizioni non possono essere integrate in uno schema esplicativo in quanto «ingranaggi che non svolgono alcuna parte nel meccanismo» (Wittgenstein, 1953, I, 271). Questa lamentela mi sembra buona, e l'unica che dobbiamo fare. (Rorty, 1982a, pp. 342-343; vedi anche Rorty, 1982b.)

Ho espresso il mio accordo, proponendo una leggera variazione della tesi «con l'incoraggiamento del Professor Rorty su ..., io sono pronto ad uscire dall'armadio e a presentarmi come un verificazionista, ma, per favore, non come un Verificazionista del Villaggio; cerchiamo di essere Verificazionisti *Urbani*» (Dennett, 1982b, p. 335). Questo libro sviluppa ulteriormente questa linea, sostenendo che se non siamo verificazionisti urbani, finiremo per tollerare ogni specie di insensatezze: epifenomenalismo, zombi, indistinguibili spettri invertiti, orsacchiotti di peluche coscienti, ragni autocoscienti.

Il punto più delicato e più rilevante per il tipo di verificazionismo che appoggio si trova nel capitolo 5, nell'argomentazione che serve a far vedere che, *non essendoci e non potendo esserci* nessuna prova a favore di un modello staliniano o di un modello orwelliano della coscienza, l'alternativa non sussiste. La confutazione ordinaria di questa asserzione verificazionista dice che io sto pregiudicando il corso della scienza; come so che nuove scoperte neuroscientifiche non possano *svelare* nuove basi per tracciare la distinzione? La replica è diretta (ma raramente udita in questi giorni): su alcuni concetti (non tutti, ma alcuni) possiamo essere sicuri di saperne abbastanza da sapere che *qualsiasi* cosa si presenterà nel cammino della scienza, non aprirà questa sorta di possibilità. Si consideri, per esempio, l'ipotesi che l'universo sia dritto e la sua negazione, l'ipotesi che l'universo sia rovesciato. Queste sono ipotesi legittime? C'è, o ci potrà mai essere, un fatto che ci possa far decidere la questione? È un peccato verificazionista dire che nessuna rivoluzione futura nella cosmologia trasformerà mai *questa* «disputa» in un fatto empirico della materia che deve essere stabilito?

Ma, in realtà, sei una sorta di comportamentista, non è vero?

Questa domanda è stata già posta; io sono contento di sottoscrivere la risposta che Wittgenstein (1953) ha dato.

307. «Non sei, in realtà, un comportamentista camuffato? In fondo non stai dicendo che all'infuori del comportamento umano tutto è finzione?» – Se parlo di una finzione, allora si tratta di una finzione *grammaticale*.
308. Come sorge il problema filosofico dei processi e stati mentali e del comportamentismo? – Il primo passo viene fatto del tutto inavvertitamente. Parliamo di processi e stati, e lasciamo indecisa la loro natura! Forse un giorno ne sapremo di più – pensiamo. Ma proprio mentre pensiamo così ci siamo impegnati per un determinato modo di considerare la cosa. Infatti abbiamo un concetto ben preciso di che cosa voglia dire: imparare a conoscere un processo più da vicino. (La mossa decisiva nel gioco di prestigio è stata fatta, ed è proprio quella che ci sembrava una mossa innocente.) – Ed ora l'analogia che avrebbe dovuto renderci comprensibili i nostri pensieri vien meno. Dunque dobbiamo negare il processo non ancora compreso, che ha luogo nel mezzo non ancora esplorato. E così sembra che abbiamo negato i processi mentali. E, naturalmente, non li vogliamo negare.

Parecchi filosofi hanno giudicato ciò che sto facendo come un rifacimento dell'attacco di Wittgenstein agli «oggetti» dell'esperienza cosciente. Infatti è così. Come 308 chiarifica, se dobbiamo evitare i giochi di prestigio, dobbiamo immaginare *prima* la «natura» degli stati e dei processi mentali. Ecco perché mi ci sono voluti nove lunghi capitoli prima di arrivare al punto in cui potessi iniziare a confrontarmi con i problemi nella loro tipica forma filosofica – che è una malformazione. Il mio debito nei confronti di Wittgenstein è enorme e di lunga data. Quando ero uno studente universitario, egli era il mio eroe, così andai a perfezionarmi ad Oxford, dove sembrava che fosse l'eroe di tutti. Quando mi accorsi che la maggior parte dei miei compagni d'università non stavano (secondo me) afferrando l'essenziale, smisi di cercare di «essere» un wittgensteiniano e mi limitai a prendere quello che pensavo di aver appreso dalle *Ricerche* e a metterlo a profitto.

APPENDICE B
(PER SCIENZIATI)

I filosofi sono spesso accusati, giustamente, di indulgere in quella che potremmo chiamare psicologia (o neuroscienza o fisica...) in poltrona e ci sono innumerevoli episodi imbarazzanti su filosofi che hanno fiduciosamente dichiarato *a priori* cose che poi sono state confutate nel laboratorio. Una reazione ragionevole di un filosofo di fronte a questo rischio appurato potrebbe essere una cauta ritirata in quelle arene concettuali in cui non c'è pericolo di dire cose che un giorno potrebbero essere confutate (o confermate) dalle scoperte empiriche. Un'altra reazione ragionevole sarebbe quella di studiare, nella propria poltrona, i frutti migliori del laboratorio, gli sforzi migliori dei teorici più legati al momento empirico, e poi procedere, con la propria filosofia, al tentativo di illuminare gli ostacoli concettuali e talvolta perfino di esporsi ai pericoli, per raggiungere in un modo o nell'altro la chiarezza sulle implicazioni di qualche idea teoretica particolare. Quando si arriva ai problemi concettuali gli scienziati non sono più immuni alle confusioni dei profani. Dopo tutto, gli scienziati passano un tempo non indifferente nelle loro poltrone, cercando di immaginare come è possibile interpretare i risultati dei loro ed altrui esperimenti, e ciò che fanno in quei momenti non è poi così differente da quello che fanno i filosofi. È un'attività rischiosa, ma corroborante.

Ecco, allora, alcune idee mezze cotte per degli esperimenti che si prefiggono di vagliare le implicazioni del modello della coscienza che ho abbozzato, esse sono tratte da un paniere molto più grande di idee – questa volta cotte solo per un quarto! – che non hanno superato le raffiche di critica dei miei pazienti informatori o che, come essi mi hanno detto, sono state già realizzate. Poiché come filosofo ho tentato di mantenere il mio modello il più generale e vago possibile, se ho svolto bene il mio lavoro, questi esperimenti dovrebbero aiutare a stabilire solo il *grado di forza* della versione del mio modello che sarà confer-

mata; se il modello dovesse risultare completamente refutato, io sarei pienamente e chiaramente smentito, e in serie difficoltà.

Sul tempo e sulla temporizzazione

Se la sequenza soggettiva è un prodotto dell'interpretazione, non una funzione diretta della sequenza effettiva, dovrebbe essere possibile creare forti effetti interpretativi di vario tipo che siano indipendenti dall'effettiva temporizzazione.

1. La passeggiata del ragno: leggeri colpetti in sequenza, a imitazione del coniglio cutaneo, ma con lo scopo di produrre giudizi illusori di *direzione*. Un semplice caso di base sarebbe costituito da due colpetti, separati spazialmente e temporalmente più o meno della stessa distanza che nel fenomeno *phi* visivo, con il compito di giudicare la direzione della «passeggiata» (il che è logicamente equivalente a giudicare la *sequenza*, ma così è un giudizio fenomenologicamente più «immediato»). Previsione: gli effetti standard del fenomeno *phi* a seconda dell'intervallo tra gli stimoli, con maggiore acuità discriminativa sulle superfici ad alta risoluzione come i polpastrelli o le labbra.

Ma ora, mentre i soggetti tengono l'indice destro e quello sinistro l'uno accanto all'altro, dare il primo colpetto su un polpastrello e il secondo sull'altro. La discriminazione della direzione dovrebbe essere molto peggiore, poiché è necessario che le comparazioni siano bilaterali. Poi si aggiunga un «aiuto» visivo; si lasci vedere ai soggetti la stimolazione delle dita, ma si fornisca un input visivo falso: il trucco nell'apparecchiatura deve far apparire una *direzione visiva* opposta alla direzione implicita nella effettiva sequenza di colpetti. Previsione: i soggetti daranno giudizi falsi ma convinti, invalidando o scartando l'informazione sulla effettiva sequenza resa disponibile dai ricettori cutanei. Se l'effetto è molto forte, può perfino invalidare i giudizi unilaterali o anche quelli sullo stesso dito che erano molto accurati in assenza di input visivo.

2. Inversione nei film. Ai soggetti viene chiesto di distinguere brevi «riprese» cinematografiche o di videotape, alcune delle quali sono state invertite o presentano sequenze manomesse o anomale. I montatori cinematografici hanno a loro disposizione molti trucchi del mestiere e una copiosa erudizione in materia, riguardo alla modificazione della sequenza dei fotogrammi del

film. Talvolta le scene vengono montate deliberatamente con i fotogrammi fuori ordine per creare effetti speciali – per accrescere l'ansietà o la sorpresa nelle scene di orrore, per esempio. Alcuni eventi sono molto rigidamente ordinati in modo naturale; ci siamo tutti divertiti nel vedere film in cui i piedi del tuffatore emergono dagli schizzi dell'acqua della piscina, seguiti poi da tutto il corpo che finisce per saltellare agile e asciutto sul trampolino. L'inversione di alcuni eventi è impercettibile – lo sventolio di una bandiera, per esempio – mentre altri si pongono in una via di mezzo; ci vorrebbe una meticolosa attenzione per decidere se il film di una palla che rimbalza stia girando regolarmente o all'indietro. Previsione: la gente non riuscirà affatto a distinguere inversioni in cui non ci siano predisposizioni interpretative – in cui una nuda sequenza deve essere rilevata e ricordata. Per esempio, tenendo la continuità del movimento e le differenze di dimensioni e forma approssimativamente costanti, i soggetti dovrebbero riuscire molto più difficilmente a distinguere (reidentificare) le sequenze che non hanno interpretazioni direzionali pregiudiziali, e distinguerle dalle loro inversioni o altre trasformazioni. (Esperimenti sulla discriminazione di melodie costituirebbero un analogo auditivo.)

3. Scrivere sul proprio piede. Un esperimento che ha lo scopo di minare i giudizi basati sull'interpretazione dei «tempi di arrivo» alla «disponibilità centrale». Supponiamo che ti si chiedesse di scrivere alcune lettere con una matita sul lato del tuo piede nudo, senza essere in grado di vedere cosa stai facendo. I segnali dai ricettori cutanei del piede «confermerebbero» che la tua azione intenzionale di scrivere una certa lettera sarebbe eseguita appropriatamente dalla matita nella tua mano. Ora si aggiunga la visione indiretta, un monitor televisivo che mostra la tua mano che scrive sul tuo piede, ma con la telecamera posta in modo tale che la tua mano impedisca di vedere la punta della matita sul piede. Questi segnali visivi aggiungerebbero ulteriori conferme dell'esecuzione delle tue azioni intenzionali. Ma ora si inserisca una breve dilazione nella televisione (uno o due fotogrammi di 33 msec ciascuno) in modo che la conferma visiva sia sempre ritardata di una quantità piccola ma costante. Prevedo che i soggetti si adatteranno facilmente a ciò. (Lo spero, perché il passo interessante è il prossimo). Dopo che si sono adattati, se la dilazione fosse immediatamente eliminata, essi avrebbero la *sensazione che la matita si fosse piegata*, perché la

percezione della traiettoria della sua punta sarebbe differita, rispetto al loro input visivo, come se la punta si trascinasse dietro la scia della sua traiettoria attesa.

4. Aggiustamento della dilazione nella giostra di Grey Walter: la continuazione dell'esperimento che misura la durata della dilazione necessaria per eliminare l'effetto di «giostra precognitiva». Prevedo che essa sarà molto più piccola dei 300-500 msec che dovremmo aspettarci da un'estensione del modello staliniano di Libet.

Sui modelli pandemonio di scelta delle parole

Come si potrebbe mostrare che «le parole vogliono essere dette»? I lapsus linguae e gli scambi di lettere possono essere controllati sperimentalmente? Gli esperimenti di Levelt finora hanno condotto a sorprendenti risultati negativi (vedi nota 2 a p. 536). Il tipo di variazione che mi piacerebbe apportare ad essi aprirebbe la possibilità di un uso «creativo» delle parole da parte dei soggetti, mentre vari materiali grezzi vengono discretamente offerti nell'ambiente affiché siano incorporati nelle loro produzioni verbali. Per esempio, i soggetti potrebbero essere preparati per l'esperimento in due differenti situazioni preliminari, in cui varie parole fuori posto, strane, vivide o piuttosto inusuali fossero lasciate cadere «casualmente» (nelle istruzioni ai soggetti, nei manifesti sui muri, ecc.); i soggetti avrebbero poi la possibilità di esprimersi su argomenti in cui queste espressioni scelte avrebbero normalmente una scarsa probabilità d'uso, di modo che le imbeccate offerte nella preparazione preliminare potrebbero mostrare di aver «acceso» le espressioni scelte e di averle messe in movimento, vagando alla ricerca dell'opportunità di essere usate. Non trovare effetti rafforzerebbe il modello di Levelt; trovarne molti (specialmente se venissero sfruttate molte «forzature») significherebbe andare a favore di un modello Pandemonio.

Esperimenti con gli inseguitori dei movimenti oculari

1. «Visione cieca» nei soggetti normali: gli esperimenti che utilizzano gli inseguitori dei movimenti oculari su soggetti normali hanno mostrato che quando uno stimolo parafoveale viene rapida-

mente spostato durante il movimento saccadico, ciò non viene notato dai soggetti (essi non riferiscono alcuna sensazione di cambiamento), ma ci sono effetti migliorativi – le latenze necessarie all'identificazione del secondo stimolo sono accorciate o no a seconda dell'informazione raccolta dallo stimolo parafoveale originale. Se ai soggetti in queste condizioni viene chiesto di provare a indovinare se lo stimolo è stato mosso oppure no (o se lo stimolo iniziale era, diciamo, una lettera maiuscola o una minuscola), essi daranno delle risposte migliori di quelle dovute semplicemente al caso? Io prevedo di sì, per una gamma interessante di domande, ma non migliore dei migliori risultati nella visione cieca.

2. *Esperimenti «con la carta da parati»*: usando un inseguitore dei movimenti oculari e variando durante i movimenti saccadici le macro e micro-caratteristiche di regioni parafoveali di superfici con «carta da parati» ripetitiva, si arrivi a stabilire la competizione necessaria per invalidare la conclusione «altre Marilyne». (Poiché i nuovi risultati di Ramachandran e Gregory mi sorprendono, io mi sbilancio a prevedere che *non* ci siano effetti *graduali* rilevabili, sebbene ai livelli in cui i soggetti si accorgono dei cambiamenti possono ben riferire strani movimenti illusori.)

3. *La scacchiera colorata*: un esperimento che ha lo scopo di mostrare quanto *limitato* sia «il pieno del campo visivo». Ai soggetti viene dato un compito di identificazione o interpretazione visiva di una scena in movimento che richiede molti movimenti saccadici: osservano figure animate in bianco e nero su uno sfondo formato da una scacchiera colorata a caso. Le caselle sono abbastanza grandi – per esempio, lo schermo CRT è diviso in una matrice di 12 x 18 quadratini colorati, a caso, in modo diverso. (I colori sono scelti casualmente di modo che la configurazione non ha un significato per il compito visivo sovrapposto allo sfondo.) È necessario che ci siano differenze di luminanza tra i quadratini, in modo che non ci sia l'effetto Liebmann, e per ogni quadratino deve essere preparato un *colore alternativo isoluminante*: un colore che, messo al posto di quello che precedentemente riempiva il quadratino, non creerebbe dei confini di luminanza radicalmente differenti ai margini (questo per tenere buoni i rilevatori dei margini di luminanza). Ora supponiamo che durante i movimenti saccadici (rilevati dall'inseguitore dei movimenti oculari) i colori della scacchiera ven-

gano cambiati; gli spettatori noterebbero che ci sono uno o più quadratini che cambiano colore molte volte in un secondo. Previsione: ci sarebbero molte situazioni in cui i soggetti non si accorgerebbero affatto che ampie porzioni dello «sfondo» stiano rapidamente cambiando colore. Perché? Perché il sistema visivo parafoveale è principalmente un sistema di allarme, formato da sentinelle incaricate di chiamare i movimenti saccadici quando viene notato un cambiamento; un tale sistema non si preoccuperebbe di registrare insignificanti colori tra le fissazioni, e quindi non gli rimarrebbe nulla con cui paragonare il nuovo colore. (Ciò dipende, naturalmente, da quanto «veloce è il film» nelle regioni che reagiscono ai colori parafoveali; ci può essere un lento periodo refrattario che annullerà l'effetto da me previsto.)

NOTE

1. *Preambolo: come sono possibili le allucinazioni?*

1. Le date fra parentesi si riferiscono alle opere elencate nella *Bibliografia*.
2. Il termine *esplosione combinatoriale* viene dall'informatica, ma il fenomeno era conosciuto molto prima dell'avvento dei calcolatori, per esempio nella favola dell'imperatore che accetta di ricompensare il contadino che gli ha salvato la vita donandogli un chicco di riso per la prima casella della scacchiera, due chicchi per la seconda, quattro per la terza, e così via, raddoppiando la quantità ad ognuna delle sessantaquattro caselle. Egli finisce per dovere allo scaltro contadino milioni di miliardi di chicchi di riso (2^{64}-1, per essere esatti). Più vicino al nostro esempio è la situazione dei romanzieri «aleatori» francesi che decisero di scrivere romanzi in cui, dopo aver letto il primo capitolo, il lettore getta in aria una monetina e legge il capitolo 2a o 2b, a seconda del risultato, e dopo di ciò legge il capitolo 3aa o 3ab o 3ba o 3bb, e così via, sempre giocando a testa o croce alla fine di ogni capitolo. Questi romanzieri ben presto si resero conto che avrebbero fatto bene a ridurre il numero dei punti di scelta se volevano evitare un'esplosione narrativa che avrebbe impedito a chiunque di poter portare l'intero «libro» a casa dalla libreria.
3. Lo sviluppo dei sistemi di «realtà virtuale», sia a scopi ricreativi che di ricerca, sta avendo attualmente un vero boom. Lo stato dell'arte è impressionante: guanti elettronicamente attrezzati che forniscono una convincente interfaccia per la «manipolazione» degli oggetti virtuali, schermi visivi montati su caschi che permettono di esplorare un ambiente virtuale di considerevole complessità. Le limitazioni di questi sistemi sono evidenti, però, e confermano la mia posizione: l'illusione può essere prodotta solo grazie a varie combinazioni di copie fisiche e di schematizzazioni (una rappresentazione a grana *relativamente* grossolana). E persino le migliori, sono esperienze di *sur*realtà virtuale, non qualcosa che può essere scambiata per la cosa reale per più di un attimo. Se vuoi davvero ingannare qualcuno e fargli credere di essere in una

gabbia con un gorilla la cosa migliore che puoi fare, ora e forse per molto tempo, è ingaggiare un attore con un costume da gorilla.
4. Per una discussione più dettagliata della questione del libero arbitrio, del controllo, della lettura della mente e dell'anticipazione, vedi il mio *Elbow Room: The Varieties of Free Will Worth Wanting*, 1984, specialmente i capitoli 3 e 4.
5. La pratica suggerisce che il gioco produce più facilmente delle buone storie se si favoriscono leggermente le risposte affermative, ponendo tra *p* e *q* la linea divisoria tra sì e no.

2. *Verso una spiegazione della coscienza*

1. Vedi il mio *Elbow Room* (1984), capitolo 4, per un'ulteriore discussione di questo tema.
2. Solo alcune anime coraggiose (sicuramente non possono aver niente da obiettare se le classifico così) sono andate controcorrente: Arthur Koestler con il suo libro provocatoriamente intitolato *Il fantasma dentro la macchina* (1967) e Popper e Eccles con il loro *L'io e il suo Cervello* (1977) sono indubbiamente degli autori eminenti; altre due difese del dualismo iconoclastiche e argutamente illuminanti sono quelle di Zeno Vendler in *Res Cogitans* (1972) e *The Matter of Minds* (1984).
3. Vedi *Murmurs in the Cathedral* (Dennett, 1989c), la mia recensione del suo libro.
4. Eccles ha proposto che la mente non fisica sia composta da milioni di «psiconi», che interagiscono con milioni di «dendroni» (tratti di cellule piramidali) nella corteccia; ogni psicone corrisponde grosso modo a ciò che Cartesio e Hume chiamerebbero un'idea – come l'idea del rosso o l'idea del cerchio o del caldo – ma a parte questa decomposizione minima, Eccles non ha nulla da dire sulle strutture, le attività, i principi attivi o le altre proprietà della mente non fisica.
5. Un'affascinante disamina dello status della teoria di Land è offerta dal filosofo C.L. Hardin in appendice al suo libro *Color for Philosophers: Unweaving the Rainbow* (1988).

3. *Una visita al giardino fenomenologico*

1. Cartesio aveva, in effetti, una concezione simile degli animali. Egli riteneva che gli animali fossero solo delle macchine molto sofisticate; come del resto anche i corpi umani, e perfino i cervelli umani. È solo grazie alla mente non meccanica e non fisica che gli esseri umani (e solo gli esseri umani) sono intelligenti e coscienti. Questa opinione era senz'altro molto sottile e in buona parte

difendibile dagli zoologi di oggi, ma era troppo rivoluzionaria per i contemporanei di Cartesio, che fecero delle caricature facilmente immaginabili delle sue idee e le trattarono irrispettosamente. Secoli dopo, queste calunnie nei confronti di Cartesio sono riprese tranquillamente da coloro che ritengono inconcepibile – o perlomeno intollerabile – la prospettiva di una spiegazione meccanicistica della coscienza. Per una trattazione illuminante, vedi Leiber (1988).

2. Questa domanda retorica implica, per alcuni, la risposta ridondante: nulla! Per esempio, McGinn (1989) argomenta la sua risposta disfattista con una disamina delle opzioni disponibili che riesce ad ignorare le possibilità che svilupperemo in capitoli successivi.
3. Perché il La sotto il Do centrale e il La sopra il Do centrale (un'ottava superiore) *hanno un suono simile*? Che cosa li rende entrambi dei La? Quale ineffabile proprietà tonale della «La-ità» hanno in comune? Certo, quando due toni qualsiasi sono separati da un'ottava (e quindi ci sembrano «uguali, solo differenti») la frequenza fondamentale dell'uno è esattamente il doppio di quella dell'altro. Il La convenzionale sotto il Do centrale ha 220 vibrazioni al secondo; il La un'ottava più in alto (il «La^3») ha 440 vibrazioni al secondo. Quando risuonano insieme, le note separate da una o più ottave sono in fase. È spiegato così il mistero di questa ineffabile relazione? «Niente affatto. Perché mai le note in fase dovrebbero suonare similmente in *questo* modo?». Beh, le note fuori fase non sembrano simili in *questa* maniera, ma possono sembrare simili in altre maniere (nel timbro, per esempio), il che ha una spiegazione differente in termini delle relazioni tra le frequenze della vibrazione prodotte. Una volta che abbiamo descritto molti modi differenti in cui le note possono sembrare simili o dissimili e li abbiamo posti in relazione alle loro proprietà fisiche e agli effetti sul nostro sistema auditivo, noi possiamo perfino prevedere, con una certa accuratezza, come delle note insolite (per esempio prodotte da un sintetizzatore elettronico) ci sembreranno. Se tutto ciò non spiega le ineffabili relazioni, che cosa bisogna ancora spiegare? (Ci occuperemo di questo sentito problema più dettagliatamente nel capitolo 12.)
4. La trattazione classica di questo tema, con ulteriori argomenti di varia qualità, è quella delle *Ricerche filosofiche* (1953) di Wittgenstein.
5. Un neurochirurgo una volta mi stava parlando di un'operazione al cervello su di un giovane affetto da epilessia. Come è consuetudine in questo tipo di operazioni, il paziente, sottoposto solo a un'anestesia locale, era completamente sveglio, mentre il chirurgo esplorava delicatamente la sua corteccia esposta, stimolando elettricamente quelle parti che si pensava di rimuovere e domandando al paziente cosa provava, per assicurarsi che non

fossero assolutamente vitali. Alcune stimolazioni provocavano lampi visivi o movimenti della mano, altre la sensazione di un confuso ronzio, ma un punto produsse una soddisfatta risposta dal paziente: «È *Outta Get Me* dei Guns N'Roses, il mio preferito gruppo *heavy metal*!».

Chiesi al neurochirurgo se avesse domandato al giovane di cantare o fischiettare al ritmo della musica che sentiva, poiché sarebbe stato affascinante apprendere quanto alta fosse la «fedeltà» della memoria così provocata. Sarebbe stata esattamente nella stessa tonalità e nello stesso tempo del disco? Quella canzone (a differenza di *Bianco Natale*) ha una versione canonica, così si sarebbe potuto semplicemente sovrapporre l'esecuzione del paziente alla registrazione standard e comparare i risultati. Sfortunatamente, benché un registratore fosse in funzione durante l'operazione, il chirurgo non fece questa richiesta al paziente. «Perché no?», chiesi; ed egli replicò: «Detesto la musica rock!».

Successivamente, nella stessa conversazione, il neurochirurgo disse incidentalmente che avrebbe operato di nuovo il giovane e io espressi la speranza che un suo tentativo per cercare di stimolare di nuovo la musica rock avesse successo per poter chiedere questa volta al giovane di cantare. «Non posso farlo», rispose il neurochirurgo, «poiché ho tolto quella parte». «Era compresa nei focolaio epilettico?», chiesi io; e egli replicò: «No, te l'ho già detto, detesto la musica rock!».

La tecnica chirurgica usata è stata sviluppata da Wilder Penfield molti anni fa ed è stata vividamente descritta nel suo *The Excitable Cortex in Conscious Man* (1958).

6. La letteratura sulla giustificazione evoluzionistica del dolore è costellata di argomenti incredibilmente miopi. Un autore sostiene che non ci può essere una spiegazione evoluzionistica del dolore perché alcuni dolori lancinanti, come quello causato dai calcoli biliari, mettono in moto un campanello d'allarme al quale nessuno poteva rispondere in modo appropriato fino agli ultimi sviluppi della medicina moderna. L'uomo delle caverne non riceveva nessun beneficio riproduttivo dal dolore dei calcoli biliari, così il dolore – almeno qualche tipo di dolore – è un mistero evoluzionistico. Ciò che questo autore ignora è il semplice fatto che per avere un sistema del dolore che accuratamente ci avverta su crisi evitabili come quelle prodotte da un artiglio o una zanna conficcati nella pancia, dobbiamo molto verosimilmente ricevere anche il premio – che solo molto più tardi si rileva tale – di un sistema che ci avverte anche su crisi che non siamo in grado di evitare. E per la stessa ragione, ci sono svariati stati interni sui quali vorremmo oggi essere avvertiti grazie al dolore (l'insorgere del cancro, per esempio), sui quali invece non sappiamo nulla presumibilmente perché il nostro passato evoluzionistico non

attribuiva nessun vantaggio per la sopravvivenza allo specifico «cablaggio» (se fosse emerso per mutazione).

7. «Che cosa penserebbe un marziano vedendo ridere un essere umano? È uno spettacolo che non può che apparirgli orribile: quei gesti convulsi, quel dimenare le membra, il torace squassato in frenetici contorcimenti». Minsky, 1985, p. 548.
8. Nell'ultima commedia di Molière, *Il malato immaginario* (1637), Argân, l'ipocondriaco del titolo, risolve finalmente i suoi problemi «diventando» egli stessò un dottore per potersi curarsi da solo. Non è richiesto lo studio, solo un po' di latino maccheronico. Egli viene messo alla prova in uno spassoso esame orale. Perché, chiede l'esaminatore, l'oppio fa dormire? Perché, replica il candidato dottore, possiede una *virtus dormitiva* – è «una polvere che fa addormentare». «*Bene, bene, bene respondere*», mormora il coro. Bella risposta! Molto informativa! Un'illuminante affermazione! Così, attualizzando un po', potremmo chiederci: «Che cosa è che rende così bella Cheryl Tiegs in fotografia?». E rispondere: «*È fotogenica*». Ah, ecco perché! (me lo chiedevo sempre). Nel capitolo 12 considereremo più dettagliatamente l'accusa di vacuità implicata nell'attribuzione di una *virtus dormitiva* a una spiegazione.

4. *Un metodo per la fenomenologia*

1. Per la macchia di colore rossa e verde, vedi Crane e Piantanida (1983) e Hardin (1988); per il confine colorato che scompare, l'effetto Liebmann (1927), vedi Spillman e Werner (1990); per la spirale acustica, vedi Shepard (1964); per l'effetto Pinocchio, vedi Lackner (1988). Per maggiori informazioni sulla prosopagnosia, vedi Damasio, Damasio e Van Hoesen (1982); Tranel e Damasio (1988); Tranel, Damasio e Damasio (1988).
2. Questo paragrafo e i successivi attingono a mie precedenti esposizioni dei punti di forza metodologici dell'eterofenomenologia: Dennett (1978c, 1982a).
3. Svariati anni fa, Wade Davis, un giovane antropologo formatosi ad Harvard, annunciò di aver scoperto il mistero degli zombi vudu; nel suo libro *The Serpent and the Rainbow* (1985) descrisse una pozione neurofarmacologica che apparentemente metteva gli esseri umani in uno stato simile alla morte; dopo essere stati sepolti vivi per vari giorni, questi individui sfortunati venivano talvolta riesumati e ridotti in uno stato di confusione e amnesia tramite la somministrazione di un allucinogeno. A causa dell'allucinogeno o del danno cerebrale provocato dalle deprivazione di ossigeno durante la permanenza nella tomba, essi si trovavano effettivamente in una condizione di totale spaesamento che li portava ad aggirarsi come fanno gli zombi nei film e talvolta

potevano essere resi schiavi. Vista la natura sensazionale delle sue affermazioni (e un film molto liberamente ispirato al suo romanzesco libro), non c'è da stupirsi se le sue scoperte si sono scontrate con un profondo scetticismo, a cui Davis ha però replicato in un secondo libro più accademico, *Passage of Darkness: The Ethnobiology of the Haitian Zombie* (1988). Vedi anche Booth (1988) e Davis (1988b).

4. In «Come cambiare idea», *Brainstorms* (1978a), ho adottato un uso convenzionale di «opinione» che mi permetteva di tracciare una distinzione tra una credenza in senso proprio e altri stati più dipendenti dal linguaggio, che chiamavo opinioni. Gli animali privi di linguaggio possono avere credenze, ma non opinioni. Le persone le hanno ambedue, ma se credi che domani sia venerdì, questa dovrebbe essere chiamata nei miei termini un'opinione, l'opinione che domani sia venerdì. Non è uno stato cognitivo in cui ti puoi trovare se sei privo di linguaggio. Sebbene non presuppongo qui una familiarità con questa distinzione, sia chiaro che le mie affermazioni si riferiscono a entrambe le categorie.

5. *Molteplici Versioni contro Teatro Cartesiano*

1. Questo ci ricorda le difficoltà che i fisici si trovano a fronteggiare quando trattano una *singolarità*, un punto in cui, proprio a causa della sua mancanza di dimensioni, varie grandezze sono infinite (date le loro definizioni). Queste difficoltà sorgono per i buchi neri, ma influenzano anche le interpretazioni di entità più terrene. Roger Penrose discute il caso che si presenta quando si cerca di applicare alle particelle le equazioni di Lorentz e quelle di Maxwell. «Ciò che le equazioni di Lorentz ci dicono di fare è di esaminare il campo elottromagnetico nel *punto* preciso in cui è situata la particella carica (e, in effetti, di provvederci di una "forza" in quel punto). Dove dev'essere localizzato quel punto se la particella ha una grandezza finita? Dobbiamo prendere il "centro" di tale particella, oppure dobbiamo fare una media del campo (per la "forza") su tutti i punti della superficie? [...] Forse il nostro compito diventerebbe più facile se supponessimo che la particella fosse *puntiforme*. Ci imbatteremmo però in altri tipi di problemi, poiché in questo caso il campo elettrico della particella diventerebbe *infinito* in prossimità della sua superficie». (Penrose, 1989, pp. 248-249)
2. Negare che la testa sia il Quartier Generale sarebbe una follia, ma non una follia senza precedenti. Philippe Pinel nel 1800 riferì il caso curioso di un uomo che cadde in «un vero delirio causato dai terrori della rivoluzione. Lo sconvolgimento della sua ragione è segnato da una particolare singolarità: egli crede di essere stato

ghigliottinato e che la sua testa sia stata gettata alla rinfusa in un mucchio insieme alle teste delle altre vittime, e, inoltre, che i giudici, pentitisi tardivamente del loro misfatto, abbiano ordinato che le teste siano riprese e riattaccate ai loro rispettivi corpi. Comunque, per un qualche errore di sorta, sulle sue spalle è stata messa la testa di un altro sventurato. L'idea che la sua testa è stata cambiata lo ossessiona giorno e notte [...] "Guarda i miei denti!", ripeteva senza sosta, "erano meravigliosi, e ora sono tutti cariati! La *mia* bocca era sana, e questa è infetta! Che differenza tra questi capelli e i capelli che avevo prima che mi cambiassero la testa!"» *Traité médico-philosophique sur l'aliénation mentale, ou la Manie.* Paris: Chez Richard, Caille et Ravier, 1800, pp. 66-7. (Ringrazio Dora Weiner per aver sottoposto alla mia attenzione questo caso affascinante.)

3. Un esempio ancor più sorprendente è costituito da un esperimento in cui il soggetto è indotto da alcuni specchi a credere ingannevolmente di star osservando la propria mano che traccia una linea, mentre in realtà sta osservando la mano di un complice dello sperimentatore. In questo caso «gli occhi vincono» così nettamente che i processi editoriali nel cervello sono indotti a concludere che la mano del soggetto venga mossa con la forza; il soggetto sostiene di sentire la «pressione» che gli impedisce di muovere la «sua» mano dove vorrebbe (Nielsen, 1963).
4. L'operazionalismo è (approssimativamente) la posizione o la linea di condotta che viene espressa da «Se non puoi scoprire la differenza, non c'è differenza» o, come spesso si sente dire, «Se schiamazza come un'anitra e cammina come un'anitra, allora è un'anitra». Per una riconsiderazione delle debolezze e dei punti di forza dell'operazionalismo, vedi Dennett (1985a).
5. C'è una regione nella corteccia chiamata MT, che risponde al movimento (e al movimento apparente). Supponiamo che una qualche attività nella MT *costituisca* la conclusione da parte del cervello che c'era un movimento concomitante. Secondo il modello delle Molteplici Versioni non ha senso chiedersi se questa conclusione sia pre-esperienziale o post-esperienziale. Sarebbe un errore chiedere, in altre parole, se l'attività nella MT sia una «reazione ad un'esperienza cosciente» (da parte di uno storiografo orwelliano) o una «decisione di rappresentare il movimento» (da parte di un montatore staliniano).
6. In realtà, Hobbes era attento ai problemi connessi a questa posizione: «Infatti, se quei colori e quei suoni fossero nei corpi o negli oggetti che ne sono la causa, non potrebbero essere separati da essi, come vediamo che accade negli specchi e negli echi prodotti dalla riflessione; in questi casi sappiamo che la cosa che vediamo è in un luogo e la sua manifestazione in un altro». (*Leviatano*, stesso capitolo). Ma questo brano si presta a varie letture molto differenti.

7. Smythies (1954). Questo pezzo eroico mostra quanto fosse difficile riflettere su questi argomenti anche solo trentasette anni fa. Egli confuta vigorosamente una versione tradizionale della teoria della proiezione e nella ricapitolazione cita approvandola la posizione analoga di Bertrand Russell: «Chiunque accetti la teoria causale della percezione è costretto a concludere che le percezioni si trovano nella nostra testa, in quanto esse giungono al termine di una catena causale di eventi fisici i quali portano dall'oggetto al cervello del percipiente. Non possiamo supporre che, alla conclusione di questo processo, l'ultimo effetto salti improvvisamente indietro al punto di partenza, come un elastico tirato quando si spezza» (Russell, 1927).
8. «È come se il nostro phaenomenista divenuto phaenomenologo dovesse, in preda alla confusione, aggrapparsi al disperato stratagemma di inventare uno spazio divino, un paradiso, in cui far risiedere il suo amato Phaenomeno; uno spazio abbastanza *reale* da soddisfare il credente che c'è in lui, ma abbastanza remoto e misterioso da nascondere Phaenomeno alla sua parte scettica. Lo spazio fenomenico è il Paradiso delle Immagini Mentali; ma se le immagini mentali si riveleranno *reali*, esse potranno comodamente risiedere nello spazio fisico nei nostri cervelli, e se si riveleranno irreali potranno risiedere, come Babbo Natale, nello spazio logico della fantasia.» Dennett (1978a), p. 294.
9. Il filosofo Jay Rosenberg mi ha fatto notare come Kant riconosca l'importanza di questo punto quando afferma che nell'esperienza il *für mich* (il «per me») e l'*an sich* (l'«in sé») sono la stessa cosa.
10. Il filosofo Ned Block una volta mi ha raccontato una sua esperienza come soggetto in un test di «lateralità». Doveva guardare fisso di fronte a sé e ogni tanto una parola (o una non parola, come GHRPE) veniva fatta lampeggiare sulla sinistra o sulla destra del suo punto di fissazione. Il suo compito era quello di premere un pulsante se lo stimolo era una parola. I suoi tempi di reazione erano nettamente più lunghi per le parole mostrate sul lato sinistro (e che entravano quindi prima nell'emisfero destro), e questo confortava l'ipotesi che lui, come la maggior parte della gente, presentava una forte lateralizzazione per il linguaggio nell'emisfero sinistro. Block non era stupito da questo risultato; ciò che lo interessava era «la fenomenologia: le parole proiettate sulla sinistra sembravano un po' sfocate». Gli chiesi se pensava che le parole fossero più difficili da identificare perché sembravano sfuocate o sembravano sfuocate perché erano più difficili da identificare. Egli ammise che non avrebbe potuto distinguere in nessun modo queste «opposte» spiegazioni causali del suo giudizio.
11. Questo modo di pensare al problema mi si è presentato per la prima volta dopo aver letto Snyder (1988), benché il suo approccio sia piuttosto differente dal mio.

1. Gli argomenti e le analisi di questo capitolo (e alcune delle discussioni del precedente) sono delle elaborazioni del materiale in Dennett e Kinsbourne (in corso di stampa).
2. Ciò non significa che il cervello non usi mai delle memorie di transito o buffer per ammortizzare l'interfaccia tra i processi interni del cervello e l'asincronia del mondo esterno. La «memoria ecoica» con la quale conserviamo le configurazioni degli stimoli per un brevissimo tempo mentre il cervello inizia ad analizzarle ne è un ovvio esempio. (Sperling, 1960; Neisser, 1967; vedi anche Newell, Rosenbloom e Laird, 1989, p. 107).
3. Mi affretto a specificare che l'abbellimento storico è mio. Francis Rawdon-Hastings, primo marchese di Hastings e secondo conte di Moira, era il governatore generale del Bengala e il comandante in capo in India nel 1815, ma non ho la più pallida idea di come e quando sia stato effettivamente informato della Battaglia di New Orleans.
4. Un «timbro» del genere può, in linea di principio, essere aggiunto al veicolo del contenuto ad ogni tappa del suo viaggio; se tutti i materiali che arrivano in un determinato luogo provenissero dallo stesso posto, tramite la stessa strada procedendo alla stessa velocità, il loro «tempo di partenza» dal luogo di origine potrebbe essere stampigliato retrospettivamente su di essi, semplicemente sottraendo una costante dal loro tempo di arrivo alla stazione intermedia. Questa è una possibilità ingegneristica che viene usata probabilmente dal cervello per realizzare certi aggiustamenti automatici per tempi di percorrenza standard.
5. Come Uttal (1979) nota, la distinzione viene ampiamente riconosciuta dai neuroscienziati: «L'essenza di gran parte delle ricerche che vengono portate avanti nel campo della codificazione sensoriale può essere distillata in un'unica importante idea – qualunque codice può rappresentare qualunque dimensione percettiva; non è necessario che ci sia una relazione isomorfa tra i dati neurali e quelli psicofisici. Lo spazio può rappresentare il tempo, il tempo può rappresentare lo spazio, il luogo può rappresentare la qualità, e certamente le funzioni non lineari neurali possono rappresentare in modo altrettanto adeguato sia funzioni psicofisiche lineari che non lineari» (p. 286). Ma sebbene questa idea sia ben conosciuta, vedremo presto come qualche teorico la intenda *fra*intendendola – reintroducendo tacitamente il non necessario «isomorfismo», immaginando un oscuro processo di traduzione o di «proiezione» che dovrebbe avvenire nella coscienza.
6. Cfr. Pylyshyn (1979, p. 278): «Nessuno [...] è disposto a parlare *letteralmente* di proprietà fisiche di un evento mentale quale il colore, le dimensioni, la massa e così via [...] sebbene diciamo che *rappresentino* (o hanno il contenuto esperienziale di) tali proprietà.

Per esempio, nessuno direbbe propriamente che un pensiero (o un'immagine) è grande o rosso, ma solo che *riguarda* una cosa grande e rossa [...] Dovrebbe apparirci curiosa, quindi, la libertà con cui parliamo di *durata* di un evento mentale».

7. Come lo psicologo Robert Efron nota: «Quando giungiamo ad osservare un oggetto con la nostra visione centrale, noi non abbiamo un'esperienza fugace di come apparirebbe in una visione periferica seguita da un'altra in una visione meno periferica. [...] Similmente, quando spostiamo la nostra attenzione da un oggetto di consapevolezza ad un altro, non c'è alcuna esperienza di "crescente" specificità del nuovo oggetto della consapevolezza – noi percepiamo il nuovo oggetto e basta» (1967, p. 721).
8. Vedi anche l'argomentazione con cui Libet rifiuta il suggerimento di MacKay di una lettura più moderata (1981, p. 195; 1985b, pag. 568). La ricapitolazione finale di Libet del 1981, d'altra parte, è inconcludente: «La mia opinione personale [...] è stata che la discrepanza temporale crea specifiche difficoltà per la teoria dell'identità, ma che esse non sono insormontabili» (p. 196). Presumibilmente esse sarebbero sicuramente insormontabili nel caso dell'interpretazione della *proiezione* all'indietro, giacché questa implica la precognizione o una causalità rovesciata o qualcosa di ugualmente misterioso e inaudito. Inoltre, Libet successivamente (1985b, p. 569) descrive queste difficoltà non insormontabili in un modo che sembra richiedere la lettura più moderata: «Benché l'ipotesi della dilazione e antecedenza non separi l'effettivo tempo dell'esperienza dal tempo della produzione neuronale, elimina però la necessità della simultaneità tra *temporalizzazione soggettiva* dell'esperienza e l'effettivo tempo (dell'orologio) dell'esperienza». Forse il supporto entusiastico di Sir John Eccles verso una interpretazione radicalmente dualistica delle scoperte ha distratto l'attenzione di Libet (e dei suoi critici) dalla tesi più moderata che egli talvolta difende.
9. In uno scritto precedente, Libet concede la possibilità di processi orwelliani e suppone che ci possa essere una significante differenza tra eventi mentali inconsci e eventi mentali consci ma effimeri: «Ci può ben essere un tipo immediato ma effimero di esperienza consapevole che non è conservata per il ricordo ai livelli coscienti dell'esperienza. Se tali esperienze esistono, comunque, il loro contenuto avrebbe una diretta significanza solo in successivi processi mentali inconsci, benché, come altre esperienze inconsce, potrebbero giocare un ruolo indiretto sulle successive esperienze coscienti» (1965, p. 78).
10. Harnad (1989) vede un insolubile problema di misurazione, ma nega proprio quello che sto affermando – che non c'è nessun momento da misurare: «L'introspezione può dirci solo quando un evento *sembra* avvenire o quale tra due eventi *sembra* avvenire prima. Non c'è un modo autonomo per confermare che la tempo-

ralizzazione reale sia effettivamente come sembrava. L'incommensurabilità è un problema metodologico, non un problema metafisico» (p. 183).

7. *L'evoluzione della coscienza*

1. Il lettore è autorizzato a supporre che ritengo giusta – o sulla strada giusta – ogni cosa che utilizzo nella mia storia, ma non deve pensare che ritengo errata qualche teoria o qualche dettaglio di una teoria dal momento che la ometto. Inoltre non deve pensare che ritengo completamente difendibile una teoria solo perché utilizzo alcuni suoi dettagli. Le stesse cose valgono anche per i miei precedenti scritti su questo argomento, dai quali attingo ampiamente.
2. John Maynard Smith è il teorico di punta, e in aggiunta al suo classico *The Evolution of Sex* (1978) ci sono vari articoli molto brillanti sui problemi concettuali nella sua raccolta di saggi *Sex, Games, and Evolution* (1989). Vedi anche R. Dawkins (1976), pp. 46-48, per una breve rassegna dei problemi.
3. L'idea di neuroni multifunzionali non è nuova, ma di recente sta guadagnando aderenti: «Sono le più o meno simultanee concatenazioni di segnali neuronali in uscita ad essere non ambigue, piuttosto che i segnali in uscita dei singoli neuroni. La convergenza di differenti concatenazioni di segnali ambigui ad ogni livello successivo risolverebbe parzialmente le ambiguità proprio come la convergenza di definizioni ambigue determina delle soluzioni uniche o quasi uniche nelle parole crociate». (Dennett, 1969, p. 83). «[...] non esiste alcuna struttura o combinazione unica di gruppi corrispondente a una data categoria o configurazione di output. Si dà invece il caso che più di una combinazione di gruppi neuronali possa fornire un particolare output, e un singolo gruppo dato possa partecipare a più di un tipo di funzione di segnalazione. Questa proprietà dei gruppi neuronali nei repertori, chiamata *degenerazione*, fornisce una base fondamentale per le capacità di generalizzazione delle mappe rientranti.» (Edelman, 1989, p. 74) Su questa caratteristica architettonica, in cui ogni nodo contribuisce a molti contenuti differenti, aveva già insistito Hebb nel suo lavoro pionieristico, *L'organizzazione del comportamento. Una teoria neuropsicologica* (1949). Essa sta alla base della «elaborazione parallelamente distribuita» o «connessionismo». Ma la funzione multipla non significa solo questo; ad un livello di analisi più macroscopico, otterremo interi sistemi che possiedono ruoli specializzati ma possono essere anche reclutati per progetti più generali.
4. L'analogia tra l'ascidia e il professore associato è stata sottolineata per la prima volta, credo, dal neuroscienziato Rodolfo Llinás.

5. Questi congegni ci fanno tornare in mente il modo, non completamente a prova d'errore, con cui Shakey distingueva le scatole dalle piramidi. Quindi Shakey non è del tutto non biologico; la biosfera è ricca di aggeggi del genere. Resta vero, però, che il sistema «visivo» di Shakey non è affatto un buon modello della visione di qualche specie animale. Questo non era ciò a cui mirava.
6. Le intuizioni fondamentali possono essere rintracciate negli scritti di Darwin e dei suoi primi divulgatori (Richards, 1987). Il neuroanatomista J. Z. Young (1965a, 1965b) è stato un pioniere della teoria selezionistica della memoria (vedi anche Young, 1979). Io ho sviluppato una versione filosofica dell'argomento fondamentale, con un abbozzo dei dettagli, nella mia dissertazione di dottorato a Oxford, 1965, che, in una versione semplificata, «Evoluzione del cervello», forma il 3 capitolo di *Contenuto e coscienza*, 1969. John Holland (1975) e altri nel campo dell'Intelligenza Artificiale hanno sviluppato degli «algoritmi genetici» per dei sistemi capaci di autoriprogettarsi o apprendere (vedi anche Holland, Holyoak, Nisbett e Thagard, 1986), e Jean-Pierre Changeux (Changeux e Danchin, 1976; Changeux e Dehaene, 1989) ha escogitato un modello neurale piuttosto dettagliato. Il neurobiologo William Calvin (1987, 1989a) offre una prospettiva differente (e più facilmente accessibile) sull'argomento nella sua teoria dell'evoluzione del cervello. Vedi anche la sua chiara e penetrante recensione (Calvin, 1989b) del *Neural Darwinism* di Gerald Edelman (1987). Più recentemente, Edelman ha pubblicato *Il presente ricordato* (1989).
7. Nella filosofia della mente questo è il problema fondamentale del *contenuto mentale o intenzionalità*, ed è noto che non c'è accordo sulle varie soluzioni proposte. La mia è data in *The Intentional Stance* (1987a).
8. Pochi intrepidi pensatori hanno sostenuto il contrario. Jerry Fodor (1975), per esempio, ha sostenuto che tutti i concetti che un individuo potrà mai avere devono essere dati fin dalla nascita e che «scattano» o «si ha accesso ad essi» solo grazie a particolari esperienze di «apprendimento». Allora Aristotele aveva il concetto di un aeroplano nel suo cervello, ed anche quello di una bicicletta – non ha mai avuto però l'occasione di usarli! A quelli che scoppiano a ridere di fronte ad un'idea così ridicola, Fodor replica che gli immunologi ridevano all'idea che la gente – Aristotele, per esempio – fosse nata con milioni di differenti anticorpi, inclusi gli specifici anticorpi per composti che sono apparsi in natura solo nel ventesimo secolo, ma che costoro ora non ridono più: si è scoperto che è proprio così. Il problema con questa idea, nella sua applicazione sia all'immunologia sia alla psicologia, sta nel fatto che le sue versioni radicali sono ovviamente false e quelle moderate sono indistinguibili dalle posizioni

a cui si oppongono. C'è una reazione combinatoria nel sistema immunitario – non *tutte* le risposte sono uno a uno tra i singoli tipi di anticorpi esistenti; analogamente, forse Aristotele aveva un concetto innato di *aeroplano*, ma aveva anche un concetto innato di *cargo jumbo jet*? E quello di *tariffa APEX andata e ritorno Boston/Londra*? Quando queste questioni saranno stabilite, in entrambi i campi, ci si accorgerà che probabilmente c'è qualcosa di simile all'apprendimento in entrambi, e qualcosa di simile ai concetti innati in entrambi.

9. Mi riferisco, naturalmente, alla teoria del significato non naturale di Paul Grice (Grice, 1957, 1969), ma per una nuova teoria della comunicazione che evita alcune delle caratteristiche più fragili e inverosimili delle teorie di Grice, vedi Sperber e Wilson (1986).
10. Che diritto ho di parlare di credenze e desideri di questi antenati non ancora completamente coscienti? La mia teoria della credenza e del desiderio, esposta in *The Intentional Stance*, fa capire che non ci sono buone ragioni per mettere tra virgolette questi termini: il comportamento degli animali «inferiori» (perfino delle rane) costituisce un dominio altrettanto adeguato per le spiegazioni svolte basandosi sull'atteggiamento intenzionale, attribuendo cioè credenze e desideri, di quello costituito dal comportamento degli esseri umani. Ma i lettori che non sono d'accordo con quella teoria sono autorizzati a pensare che questi termini siano usati qui in un senso metaforicamente esteso.
11. Sulla comunicazione tra primati e l'ancora irrisolta questione empirica della possibilità o meno che le scimmie e i gorilla siano capaci di inganno deliberato, vedi Dennett (1983, 1988c, 1988d, 1989a); Byrne e Whiten (1988); Whiten e Byrne (1988).
12. Nel *Giardino dei sentieri che si biforcano*, Jorge Luis Borges escogita una versione diabolicamente intelligente di questa strategia, che mi astengo dal descrivere, perché non voglio sciupare uno squisito finale.
13. Questo paragrafo è tratto dal mio *Memes and the Exploitation of Imagination* (1990a).
14. Per una buona disamina contemporanea delle controversie riscontrabili nella letteratura sull'evoluzione del linguaggio, vedi Pinker e Bloom (1990) e i successivi commenti.
15. I puristi potrebbero obiettare che io uso il termine *macchina virtuale* in un senso più ampio di quello che essi raccomandano nell'informatica. Io rispondo che, come Madre Natura, quando vedo un comodo termine da «estrapolare utilmente» e di cui servirmi per un uso esteso (Gould, 1980), di solito non ho remore.
16. Oppure potrebbe non essere affatto una macchina virtuale. Potrebbe essere una macchina reale appositamente cablata in modo rigido per degli scopi particolari, come una macchina

Lisp, che deriva dalle macchine *virtuali* Lisp e che è progettata fin nei suoi chip al silicio per far girare il linguaggio di programmazione Lisp.

17. I «neuroni logici» di McCulloch e Pitts (1943) furono in realtà escogitati contemporaneamente all'invenzione del calcolatore seriale e influenzarono il pensiero di von Neumann, e condussero poi ai percettroni degli anni Cinquanta, gli antenati del connessionismo odierno. Per una breve illustrazione storica vedi Papert (1988).
18. Per un approfondimento sulle implicazioni della velocità del mondo reale e le sue implicazioni per l'Intelligenza Artificiale, vedi «Fast Thinking» nel mio *The Intentional Stance* (1987a).
19. Per un'interessante discussione sul disaccordo (apparente) tra due scuole di pensiero nell'Intelligenza Artificiale, ragionamento opposto a ricerca, vedi Simon e Kaplan, 1989, pp. 18-19.

8. *Come le parole collaborano con noi*

1. Dan Sperber e Deirdre Wilson (1986) aprono una nuova prospettiva sul modo in cui componiamo le nostre comunicazioni insistendo su modelli del come le cose effettivamente *funzionano*, in colui che parla e colui che ascolta; in questo modo si differenziano da una pratica recente e diffusa tra filosofi e linguisti, e che li spinge a trascurare i meccanismi quando si interessano alle ricostruzioni razionali dei presunti compiti e i loro requisiti. Ciò permette a Sperber e Wilson di sollevare delle considerazioni di praticabilità ed efficienza: principi del minimo sforzo, e attenzione alla temporalizzazione e alla probabilità. Essi mostrano poi da questa nuova prospettiva come alcuni «problemi» tradizionali scompaiono – in particolare, il problema di come un ascoltatore trova la «giusta» interpretazione di quello che il parlante intendeva. Sebbene non spingano il loro modello al livello del processo evoluzionistico del tipo che abbiamo appena considerato, invitano certamente a fare una tale elaborazione.
2. Come nota Levelt: «Se si potesse mostrare, per esempio, che la generazione del messaggio è direttamente influenzata dall'accessibilità a lemmi o parole, si avrebbe la prova di una retroazione diretta tra il Formulatore e il Concettualizzatore. Questo è un problema empirico ed è possibile sottoporlo ad esame. [...] Finora i tentativi di dimostrare una tale retroazione hanno dato esito negativo» (p. 16). I tentativi a cui egli si riferisce sono degli esperimenti strettamente controllati in cui al parlante è dato un compito molto specifico: come *descrivere l'immagine sullo schermo il più rapidamente possibile* (pp. 276-282). Queste sono eccellenti prove negative – io stesso sono stato sorpreso della mancanza di effetti in questi esperimenti – ma, come egli riconosce, non sono affatto

conclusive. Non è realmente *ad hoc* affermare che l'artificialità di queste situazioni sperimentali ha spazzato via la dimensione opportunistica/creativa dell'uso linguistico. Ma forse Levelt ha ragione; forse l'unica retroazione esistente tra il Formulatore e il Concettualizzatore è *indiretta*: il tipo di retroazione che una persona può generare *solo* parlando esplicitamente con se stessa e poi formandosi un'opinione su ciò che si scopre a dire.

3. Levelt mi dice che lui stesso è un inveterato cacciatore di doppi sensi (in olandese, la sua madrelingua) e che sa bene come fa: «Con un allenamento sviluppato nel corso di tutta la vita, giro e rigiro quasi ogni parola che ascolto. Poi controllo (piuttosto consciamente) il significato del risultato. Nel 99,9 per cento dei casi non ne viene fuori nulla di divertente. Ma l'un per mille riesce bene e subito lo dico» (comunicazione personale). Questo è un esempio perfetto di una risoluzione di problemi in stile neumanniano: seriale, controllato – *e cosciente*! La questione è se ci sono altri modi, più pandemonici, di generare inconsciamente le arguzie.
4. Secondo *The Oxford Dictionary of Quotations* (seconda edizione, 1953), questa frase famosa è attribuita anche a Phineas T. Barnum. Poiché Barnum è un illustre ex-studente e benefattore generoso della mia università, mi sento obbligato ad attirare l'attenzione sulla possibilità che Lincoln non sia la fonte originaria di questo meme altamente replicativo.
5. Questo ci ricorda l'opinione di Freud sul modo in cui funziona il «preconscio»: «Allora la domanda "Com'è che qualche cosa diventa cosciente?" andrebbe formulata più adeguatamente nel modo seguente: "Com'è che qualche cosa diventa preconscio?". La risposta dovrebbe essere: "Attraverso il collegamento con le rispettive rappresentazioni verbali"» (*L'Io e l'Es*, 1922; ed. it. Boringhieri, vol. 9, p. 483).
6. Levelt mi dice che alcune ricerche in corso all'Istituto di psicolinguistica «Max Planck» di Nijmegen gettano dubbi su questa opinione comunemente accettata. Il lavoro di Heeschen suggerisce che ad un qualche livello gli ammalati di afasia gergale o di Wernicke rivelano ansie sulle loro deficienze e sembra che adottino una strategia di ripetizione nella speranza di riuscire a comunicare.
7. Un altro fenomeno linguistico anomalo è costituito dal familiare sintomo della schizofrenia, «l'ascoltar voci». È ora assodato in modo piuttosto certo che la voce che lo schizofrenico «ode» è la propria; egli sta parlando silenziosamente a se stesso senza accorgersene. Basta un accorgimento così semplice come quello di far tenere la bocca ben aperta al paziente per interrompere le voci (Bick e Kinsbourne, 1987). Vedi anche Hoffmann (1986) e i commenti di Akins e Dennett, *Who May I Say Is Calling?* (1986).

1. Edelman (1989) è un pensatore che ha tentato di mettere tutto insieme, dai dettagli della neuroanatomia alla psicologia cognitiva, dai modelli computazionali alle più astruse controversie filosofiche. Ne è risultato un istruttivo fallimento. Esso mostra molto dettagliatamente come ci siano innumerevoli domande molto differenti che devono ricevere una risposta prima di poter affermare di aver raggiunto una completa teoria della coscienza, ma mostra anche come nessun teorico possa apprezzare tutte le sottigliezze dei problemi sollevati dalle varie prospettive. Edelman ha interpretato erroneamente, e poi rapidamente liquidato, il lavoro di molti suoi potenziali alleati, e così ha isolato la sua teoria dal tipo di attenzione benevola e dotta di cui ha bisogno per porla a riparo dai suoi errori e imperfezioni. Ciò fa sorgere la possibilità analoga che anch'io abbia sottostimato alcuni dei lavori verso i quali esprimo disaccordo in queste pagine; senza dubbio l'ho fatto, e spero che coloro di cui ho mal interpretato i prodotti intellettuali si facciano avanti e provino (di nuovo) a spiegare quel che non ho afferrato.
2. I funzionalisti hanno preso un'abitudine «riquadrologica» – disegnare diagrammi in cui le varie componenti funzionali sono sistemate in riquadri separati, pur negando esplicitamente che questi riquadri abbiano un qualche significato anatomico. (Io stesso sono colpevole di indulgere in questa pratica e incoraggiarla; vedi le figure ai capitoli 7, 9 e 11 di *Brainstorms*.) Penso ancora che «in linea di principio» questa sia una buona tattica, ma in pratica tende a non far vedere al funzionalista la possibilità di scomposizioni alternative e soprattutto di funzioni multiple sovrapposte. L'immagine della separazione spaziale tra memoria di lavoro e memoria a lungo termine – un'immagine vecchia quanto la colombaia di Platone – gioca un ruolo tutt'altro che banale nel modo in cui i teorici analizzano il compito del riconoscimento. Un esempio sorprendente: «Il bisogno di simboli sorge perché non è possibile che tutta la struttura implicata nella computazione sia assemblata in anticipo nel luogo fisico della computazione. Così è necessario rivolgersi ad altre parti (distali) della memoria per ottenere la struttura addizionale». (Newell, Rosenbloom e Laird, 1989, p. 105). Ciò conduce in modo piuttosto diretto all'immagine di *simboli mobili*, e poi (in quelli che sono acriticamente attratti da questa immagine), allo scetticismo su tutte le architetture connessioniste, sulla base del fatto che gli elementi che assomigliano di più ai simboli in tali architetture – i nodi che in un modo o nell'altro ancorano la semantica al sistema – sono immobili nella loro ragnatela di interconnessioni. Vedi, per es., Fodor e Pylyshyn (1988). Questa contrapposizione tra elementi semantici fissi e elementi semantici mobili è un modo di

guardare a un problema ancora irrisolto nella scienza cognitiva. Probabilmente non è un buon modo di vederlo, ma non scomparirà fino a che non sarà sostituito da una prospettiva migliore, ancorata a una convinta accettazione – opposta a un'isterica liquidazione – dei fatti fondamentali della neuroanatomia funzionale.

3. In altre parole, le attrattive di una «teoria causale del riferimento» sono ugualmente evidenti sia per gli scienziati cognitivi sia per i filosofi.
4. Fodor nota una variante di questo problema nella sua discussione sul «pensare un concetto» (1990, pp. 80-81).
5. Le teorie dell'attenzione che si rifanno alla metafora del riflettore sono state popolari per anni. Le versioni più grossolane fanno l'errore di supporre troppo letteralmente che ciò che viene di volta in volta illuminato o potenziato dal riflettore sia una zona dello *spazio visivo* – proprio come la luce della ribalta in un teatro che illumina una zona del palcoscenico alla volta. Le versioni più difendibili – ma anche a questo punto più impressionistiche – specificano che è una porzione dello *spazio concettuale o semantico* ad essere potenziato (immagina, se puoi, un fascio di luce nel teatro che colpisce solo i *Capuleti* o tutti e solo gli *amanti*). Vedi Allport (1989) sulle difficoltà delle teorie del riflettore.
6. Strettamente correlate a questa distinzione sono la mia trinità di atteggiamento intenzionale, atteggiamento del progetto e atteggiamento fisico (Dennett, 1971), e l'identificazione del «livello della conoscenza» al di sopra del «livello del sistema fisico dei simboli» da parte di Allen Newell (1982). Vedi Dennett (1987a, 1988e) e Newell (1988).
7. Come Marr nota: «Diventa possibile, separando le spiegazioni in livelli differenti, fare affermazioni esplicite su ciò che viene computato e perché, e costruire teorie che affermano che ciò che viene computato è ottimale in un certo senso o che è garantito un funzionamento corretto» (p. 19). Per saperne di più sui benefici e le trappole di questa ingegneria alla rovescia vedi Dennett (1971, 1983, 1987a, 1988d); Ramachandran (1985).
8. In *Minimal Rationality* (1986), il filosofo Christopher Cherniak analizza le prospettive e i limiti dei processi deduttivi del tipo reso possibile dalla creazione di questo forum aperto. Vedi anche Stalnaker (1984).
9. Jackendoff (1987) adotta una tattica leggermente differente. Egli divide il problema mente-corpo in due e cerca di spiegare con la sua teoria come la mente *computazionale* si adatti al corpo; questo lo lascia con un irrisolto «problema mente-mente» – qual è la relazione tra la mente fenomenologica e la mente computazionale? Invece di concedere questo mistero

residuo, io mi ripropongo di far vedere come il modello delle Molteplici Versioni, in concerto con il metodo dell'eterofenomenologia, dissolva entrambi i problemi in una volta.

10. *Mostrare e dire*

1. Questa utile semplificazione è proprio una delle molte sottigliezze che sono state successivamente indagate dai ricercatori; oggigiorno disponiamo di buone prove a favore dell'esistenza di un effetto di «inerzia» o «momento» nella trasformazione di immagini. Vedi Freyd (1989).
2. L'animazione stupefacente, ma chiaramente poco fluida, nel popolare programma per PC-IBM *Flight Simulator*, mostra i limiti dell'animazione in tempo reale di scene piuttosto complesse con un piccolo calcolatore.
3. Chiamo questo congegno un *Vorsetzer* perché mi ricorda lo straordinario piano meccanico tedesco che portava questo nome; esso comprendeva un'unità separata con ottantotto «dita» che poteva «sedere di fronte» ad un normale pianoforte e premere i tasti dall'esterno proprio come un pianista umano. (È importante osservare che questo congegno è un *Vorsetzer* – un sedente di fronte – ma non un *Vorsitzende* – un presidente!)
4. Una volta che abbiamo le etichette, possiamo indicare *qualunque* proprietà dell'oggetto, non solo le proprietà spaziali o le proprietà visibili – come nelle facezie dei vecchi libri da colorare: «Ecco il mio capo. Coloralo odioso».
5. Vedi Kosslyn (1980) per una discussione sul formato. Jackendoff (1987) offre un'analisi affine di ciò che chiama la forma delle strutture informative.
6. Kosslyn (1980) fornisce non solo un'argomentazione dettagliata in difesa dell'insieme particolare delle sue risposte (a quel tempo), ma anche un'eccellente rassegna dei lavori sperimentali e teoretici degli altri sull'immaginazione. Una buona rassegna del lavoro nel decennio successivo può essere trovata in Farah (1988) e in Finke, Pinker e Farah (1989).
7. Come dice Marvin Minsky, «Non vi è nulla di strano nell'idea che si possano percepire gli eventi interni al cervello. Gli agenti sono agenti e per un agente è altrettanto facile essere cablato in modo da rilevare *un evento cerebrale causato dal cervello*, quanto esserlo in modo da rilevare *un evento cerebrale causato dal mondo.* (1985, p. 290).
8. Nel capitolo 7 (p. 251), avevo chiesto «pungolati alla superficie di *che*?» e avevo promesso di rispondere successivamente. Ecco la mia risposta. La superficie (metaforica) è determinata dal formato delle interazioni tra le parti.
9. È interessante confrontare tracce differenti di quest'idea dell'«U-

tente nel cervello» nel lavoro di vari pensatori molto disparati. Ecco Minsky (1985): «Esagerando un po', si può dire che ciò che chiamiamo "coscienza" consiste in poco più che una serie di menu che appaiono di quando in quando su schermi video mentali usati da altri sistemi [p. 103] [...] Dividiamo il cervello in due parti, A e B; colleghiamo gli ingressi e le uscite del cervello A al mondo reale, in modo che esso possa percepire ciò che vi accade. Colleghiamo invece il cervello B non al mondo esterno, bensì in modo tale che il cervello A sia il mondo del cervello B! [p. 106]». Minsky si rifiuta saggiamente di arrischiarsi a tracciare qualsiasi linea divisoria anatomica, ma altri sono pronti a rischiare. Quando Kosslyn avanzò per la prima volta l'ipotesi che la coscienza fosse una macchina virtuale era incline a localizzare l'Utente nei lobi frontali (vedi anche Kosslyn, 1980, p. 21); e più recentemente Edelman ha seguito la propria argomentazione fino alla stessa conclusione, messa in termini di «memoria del sé e non sé dominata dai valori» che egli colloca nei lobi frontali e a cui assegna il compito di interpretare le produzioni del resto del cervello (Edelman, 1989, p. 127ss).

10. Ciò è perlomeno strettamente affine a un tema centrale nell'ultimo lavoro di Wittgenstein, ma Wittgenstein evitò di sviluppare qualunque spiegazione o modello preciso della relazione tra quello che diciamo e ciò di cui parliamo quando riferiamo (apparentemente) i nostri stati mentali. La filosofa Elizabeth Anscombe, nel suo classico, frustrante ed oscuro, *Intention* (1957), prova a colmare questa lacuna lasciata da Wittgenstein, sostenendo che è errato affermare che *sappiamo* quali siano le nostre intenzioni; piuttosto possiamo solo *dire* quali sono le nostre intenzioni. Ella tentò anche di caratterizzare una categoria di cose che potessimo *conoscere senza osservazione*. Un'imperfetta discussione di queste affermazioni è contenuta nel mio *Contenuto e coscienza* (1966), capitoli 8 e 9. Ho sempre pensato che ci fosse qualcosa di giusto, importante e originale, latente in questi passi. Il mio secondo passaggio su di essi può essere trovato in «Verso una teoria cognitiva della coscienza» (1978) ristampato in *Brainstorms*, specialmente i paragrafi IV e V (pp. 265-274 di *Brainstorms*). Questo paragrafo rappresenta il mio tentativo attuale di portare alla luce queste idee e si discosta sostanzialmente da entrambe le versioni precedenti.

11. In *Brainstorms*, sfruttai questa caratteristica della psicologia del senso comune nella mia discussione su «la varietà β di credenze fenomenologiche» (1978a p. 281 sgg.).

11. *Smantellare il programma di protezione del testimone*

1. Tim Shallice, in *From Neuropsychology to Mental Structure* (1988),

fornisce una discussione molto dettagliata e aggiornata sui ragionamenti implicati nell'analisi di questi esperimenti naturali. Parecchi libri recenti forniscono buone spiegazioni divulgative di alcuni di questi casi affascinanti: Howard Gardner, *The Shattered Mind* (1975) e Oliver Sacks, *L'uomo che scambiò sua moglie per un cappello* (1985).

2. Si noti che i dettagli del danno neurologico di per sé (senza le negazioni) non proverebbero nulla; è solo mettendo in corrispondenza il danno neurologico con i resoconti (credibili) e le prove comportamentali che otteniamo qualche ipotesi su quali parti del cervello siano essenziali per un certo fenomeno cosciente.
3. Senza la conferma delle analisi cerebrali che mostrano il danno corticale, ci sarebbe sicuramente anche un diffuso scetticismo sull'autenticità degli scotomi dei soggetti con visione cieca. Vedi, per es., Campion, Latto e Smith (1983) e Weiskrantz (1988).
4. Il filosofo Colin McGinn (1991) dice di un'immaginaria paziente con visione cieca: «Dal punto di vista del comportamento, può funzionare in modo molto simile a quello di una persona vedente; fenomenologicamente, si considera cieca» (p. 111). Questo è semplicemente falso; lei non può affatto funzionare comportamentalmente come una persona vedente. McGinn seguita a sottovalutare la sua sorprendente affermazione: «Inoltre, facciamo gli ingenui per un attimo, non sembra che i pazienti con visione cieca abbiano effettivamente delle esperienze visive quando effettuano le loro sorprendenti discriminazioni? [...] Non assomigliano alla gente normale quando non c'è *nulla* di esperienziale in atto» (p. 112). Di nuovo, questo è falso. Sembra infatti che *non* abbiano esperienze visive, *perché devono essere imbeccati*. Se non dovessero essere sollecitati, sembrerebbe effettivamente che abbiano esperienze visive – al punto tale che non crederemmo alle loro affermazioni di cecità!
5. «Se potesse osservare la sua risposta galvanica cutanea, sarebbe in una forma migliore» – Larry Weiskrantz, parlando di uno dei suoi pazienti con visione cieca, ZIF, Bielefeld, Maggio 1990.
6. La mia risposta a questa domanda è il mio libro *The Intentional Stance* (1987).
7. La sua identificazione del ditale è un *effetto* conseguente al suo diventare cosciente di esso oppure è una causa precedente del suo diventarne cosciente? Questa è una domanda – orwelliana o staliniana? – che il modello delle Molteplici Versioni ci insegna a non fare.
8. In condizioni normali, la localizzazione (lo «scovarlo») e l'identificazione vanno a braccetto; scovare «la cosa che deve essere identificata» è una precondizione per identificarla. Ma questa normale coincidenza maschera un fatto sorprendente: il meccanismo di identificazione e quello di localizzazione sono caratterizzati da un ampio grado di indipendenza nel cervello, sono

localizzati in regioni differenti della corteccia (Mishkin, Ungerleider e Macko, 1983), ed è quindi possibile bloccare uno senza danneggiare l'altro. Esistono rare patologie in cui il soggetto può prontamente identificare cosa vede senza essere affatto capace di localizzarlo nello spazio personale, e patologie complementari nelle quali i soggetti possono localizzare uno stimolo visivo – indicarlo, per esempio – ed avere ciononostante grandi difficoltà a identificare l'oggetto, malgrado abbiano una visione altrimenti piuttosto normale. La psicologa Anne Treisman (1988; Treisman e Gelade, 1980; Treisman e Sato, 1990; Treisman e Souther, 1985) ha condotto un'importante serie di esperimenti a favore della sua tesi che *vedere* dovrebbe essere distinto dall'*identificare*. Quando qualcosa viene visto, secondo il suo modello, il cervello compone una «replica» dell'oggetto. Le repliche sono «autonome rappresentazioni episodiche temporanee» – e la loro creazione è il preambolo per l'ulteriore identificazione, qualcosa che viene realizzato tramite una ricerca nella propria memoria semantica usando un processo sul tipo di quello dei modelli dei sistemi di produzione. Una replica non deve essere definita da una precisa localizzazione nello spazio personale, comunque, se capisco il suo modello, e per queste ragioni non è impossibile che soggetti nello stato di Betsy (prima di trovare il ditale) possano ottenere risultati migliori del caso se sollecitati a fare le loro congetture a scelta forzata sulla presenza o l'assenza attuale del ditale dal loro campo visivo. Per esperimenti a questo proposito, vedi Pollatsek, Rayner e Henderson (1990).

9. Per esempio, la latenza nella risposta ad alcuni di questi compiti percettivi, anche in soggetti allenati, è piuttosto alta – da otto a quindici secondi per varie identificazioni molto semplici, per esempio (Bach-y-Rita, pag. 103). Questo dimostra che il flusso d'informazione nella visione protesica è estremamente lento, se paragonato alla vista normale.

10. Baud è l'unità di misura standard per la velocità del flusso dell'informazione digitale (significa approssimativamente, bit al secondo). Per esempio, se il tuo calcolatore comunica tramite le linee telefoniche con altri calcolatori, può trasmettere le sue stringhe di bit a 1.200 baud o 2.400 baud o a una velocità superiore. È necessaria una velocità in baud circa quattro volte superiore per trasmettere immagini animate ad alta risoluzione in tempo reale – un chiaro esempio in cui un'immagine vale in effetti più di mille parole. I normali segnali televisivi sono analogici, come un disco fonografico, invece che digitali, come un compact disc, così la velocità del suo flusso d'informazione è misurato in larghezza di banda, piuttosto che in baud. Il termine precede i calcolatori; il codice Baudot, così chiamato per il suo inventore (come per il codice Morse), era il codice telegrafico internazionale standard adottato nel 1880, e il baud rappresentava il

numero di elementi del codice trasmessi al secondo. Dicendo «velocità in baud» invece di «larghezza di banda», non intendo implicare che è meglio concepire in termini digitali il trattamento dell'informazione nel cervello.

11. Esistono altre specie di algoritmi di compressione che non si basano su una simile suddivisione dell'immagine in regioni ugualmente colorate, ma non me ne occuperò ora.
12. Altre creature hanno differenti solidi dei colori – o ipersolidi! Noi siamo «tricromatici»: abbiamo tre tipi differenti di cellule trasduttrici fotopigmentate nei coni delle nostre retine. Altre specie, come i piccioni, sono tetracromatici; il loro spazio colorato soggettivo dovrebbe essere rappresentato, numericamente, come un iperspazio a quattro dimensioni. Altre specie ancora sono dicromatiche, tutte quelle la cui discriminazione del colore potrebbe essere rappresentata su un singolo piano bidimensionale. (Si noti che il «bianco e nero» è uno schema rappresentazionale ad una sola dimensione, con tutti i grigi possibili rappresentabili come distanze differenti su una linea tra 0 e 1). Per delle riflessioni sulle implicazioni di questa incommensurabilità dei sistemi di colore, vedi Hardin (1988) e Thompson, Palacios e Varela (in corso di stampa).
13. Come se avessero aderito a questo suggerimento, V. S. Ramachandran e R. L. Gregory (sottoposto) hanno appena eseguito alcuni esperimenti su ciò che loro chiamano (credo, in modo fuorviante) scotomi indotti artificialmente, nei quali ci sono buone prove a favore di un riempimento graduale di *texture* e dettagli. C'è una differenza fondamentale tra le loro circostanze sperimentali e le condizioni che sto descrivendo: nei loro esperimenti c'è una competizione tra due fonti d'informazione, e una viene (gradualmente) sopraffatta. Il fenomeno del riempimento spaziale graduale delle *texture* è una scoperta importante, ma non ci porta oltre un modello analogo alla figura 11.11. Prima che si possa dare una interpretazione definitiva a questi esperimenti si devono risolvere ulteriori questioni su di essi.
14. Per esempio, gli esperimenti iniziali di Roger Shepard sulla rotazione mentale di disegni di cubi hanno mostrato che certamente ai soggetti *sembrava* di avere a che fare con rappresentazioni rotanti approssimativamente continue delle forme che stavano immaginando, ma sono serviti ulteriori esperimenti, sulle effettive proprietà temporali delle sottostanti rappresentazioni, per fornire una (parziale) conferma dell'ipotesi che essi stavano realmente facendo quello che sembrava loro di fare. (Vedi Shepard e Cooper, 1982.)
15. Nell'Appendice B, suggerirò alcuni «esperimenti con la carta da parati» che potrebbero mettere alla prova questa tesi empirica.
16. Per un'opinione differente, vedi Bisiach ed altri (1986) e McGlynn e Schacter (1989), i cui modelli di anosognosia sono

simili, ma si rimettono alla «riquadrologia» di *sistemi* separati, specialmente quello di McGlynn e Schacter, che postula un *sistema di consapevolezza cosciente*, che riceve *input* dai moduli.

12. *Quali qualia?*

1. Variazioni sullo stesso tema possono essere trovate in Humphrey (1976, 1983a) e in Thompson, Palacios e Varela (in stampa).
2. I filosofi sono oggi attratti dal concetto di *razze naturali*, reintrodotto nella filosofia da Quine (1969), il quale può rimpiangere ora il modo in cui esso è diventato un sostituto per il concetto, dubbio ma in fondo popolare, di *essenze*. «Le cose verdi, o almeno gli smeraldi verdi, sono una razza», osserva Quine (p. 116), manifestando il proprio riconoscimento del fatto che mentre gli smeraldi possono essere una razza naturale, le cose *verdi* probabilmente non lo sono. La discussione attuale intende prevenire uno degli errori più allettanti del naturalismo in poltrona: la presupposizione che qualunque cosa faccia la natura sia una razza naturale. I colori *non* sono «razze naturali» proprio *perché* sono il prodotto dell'evoluzione biologica, che ha una tolleranza nei confronti dei confini incerti quando crea categorie che orrificherebbero qualunque filosofo con la passione per le definizioni chiare e nette. Se la vita di qualche creatura dipendesse dall'ammassare insieme la luna, il gorgonzola e le biciclette, puoi stare quasi sicuro che Madre Natura troverebbe il modo per fare «vedere» a lui queste cose come «intuitivamente la stessa razza di cose».
3. La primatologa Sue Savage-Rumbaugh mi ha informato che i bonobi, o scimpanzé pigmei, allevati in laboratorio, non mostrano segni di disgusto innato nei confronti dei serpenti, a differenza degli scimpanzé.
4. La subitaneità sarebbe importante, perché se avvenisse molto gradualmente, potresti non essere in grado di accorgertene. Come Hardin (1990) ha sottolineato, l'ingiallimento graduale del cristallino con l'età, sposta lentamente la tua percezione dei colori primari; se ci viene mostrata una ruota di colori e ci viene chiesto di indicare il rosso puro (il rosso senza tracce di arancione o viola), il punto del continuum che indichiamo è in parte funzione dell'età.
5. «Il fatto che diciamo così volentieri: "L'importante è *questo*" – accennando, rivolti a noi stessi, alla sensazione, – mostra già quanto forte sia in noi l'inclinazione a dire qualcosa che non è una comunicazione». Wittgenstein (1953, I, 298).
6. Sarebbe un atto di disperata disonestà intellettuale citare quest'asserzione fuori del contesto!
7. Cfr. Lockwood (1989): «Come sarebbe apparsa la coscienza se

fosse apparsa come miliardi di minuscoli atomi che si agitano sul posto?» (pp. 15-16).

8. Ho presentato le idee principali nelle mie riflessioni su Borges, nell'*Io della mente* (Hofstadter e Dennett, 1981, pp. 337-340) e le ho riunite in un discorso presentato al Houston Symposium nel 1983, *The Self as the Center of Narrative Gravity*. Aspettando che apparisse il volume di quel simposio, ho pubblicato una versione piuttosto ridotta del mio discorso sul «Times Literary Supplement», 16-22 sett. 1988, con il titolo noioso – non mio – *Why everyone is a novelist*. La versione originale, con il titolo *The Self as the Center of Narrative Gravity*, è ancora in corso di pubblicazione in F. Kessel, P. Cole e D. Johnson (a cura di), *Self and Consciousness: Multiple Perspectives*, Hillsdale, NJ: Erlbaum.

13. *La realtà dei sé*

1. È interessante notare che Nagel, nel 1971, si stava già ponendo esplicitamente questa domanda (p. 398), prima di rivolgere la sua attenzione ai pipistrelli – un argomento che discuteremo nel prossimo capitolo.
2. E come sappiamo che *noi* stiamo facendo qualcosa? Da dove otteniamo quel poco di auto-conoscenza che usiamo come fulcro su cui basarci? Ad alcuni filosofi (Castañeda, 1967, 1968; Lewis, 1979; Perry, 1979), questa è sembrata una questione fondamentale e ha generato una letteratura di una complicazione incomparabile. Se questo è un problema filosofico sostanziale, deve esserci qualcosa di sbagliato nella risposta «banale» (ma io non capisco perché): noi otteniamo la nostra auto-conoscenza basilare e originaria nello stesso modo dell'aragosta; siamo semplicemente cablati in quella maniera.

14. *La coscienza nella nostra immaginazione*

1. Prova ad immaginare cosa doveva pensare Ji Hu-Min, un mio studente laureato all'università di Pechino, che come introduzione alla filosofia anglo-americana della mente (mentre il suo inglese era ancora abbastanza rudimentale) partecipava ad un seminario in cui gli studenti e i professori dibattevano calorosamente che cosa sarebbe successo se l'intera popolazione cinese fosse in qualche modo forzata a partecipare ad una gigantesca realizzazione di un programma di IA ritenuto cosciente (l'esempio di Block) e poi passavano a discutere, con la stessa mancanza di sensibilità nei confronti di un osservatore cinese, l'esempio della Stanza Cinese di Searle.
2. La confutazione definitiva, che non ha ancora ricevuto una

risposta adeguata da parte di Searle, è quella di Douglas Hofstadter, in Hofstadter e Dennett (1981), pp. 360-369. Ci sono state altre critiche incisive nel corso degli anni. In *Fast Thinking* (in Dennett, 1987a), ho offerto una nuova diagnosi delle fonti di ambiguità del suo esperimento ideale. La sua risposta è stata di dichiarare, senza nessuna argomentazione, che tutte le mie osservazioni erano irrilevanti (Searle, 1988b). A nessun prestigiatore piace che i suoi trucchi vengano svelati in pubblico.

BIBLIOGRAFIA

Akins, K. A. (1989), *On Piranhas, Narcissism and Mental Representation: An Essay on Intentionality and Naturalism.* Tesi di dottorato, Dipartimento di Filosofia, University of Michigan, Ann Arbor.

— (1990), *Science and Our Inner Lives: Birds of Prey, Bats, and the Common (Featherless) Biped* in *Interpretation and Explanation in the Study of Animal Behavior*, Vol. I, a cura di M. Bekoff e D. Jamieson, Westview, Boulder, pp. 414-427.

Akins, K. A. e Dennett, D. C. (1986), *Who May I Say Is Calling?*, «Behavioral and Brain Sciences», 9, pp. 517-518.

Allman, J., Meizin, F. e McGuinness, E. L. (1985), *Direction-and Velocity-Specific Responses from beyond the Classical Receptive Field in the Middle Temporal Visual Area*, «Perception», 14, pp.105-126.

Allport, A. (1988), *What Concept of Consciousness?* in Marcel e Bisiach (a cura di) (1988), pp. 159-182.

— (1989), *Visual Attention* in *Foundations of Cognitive Psychology*, a cura di M. Posner, MIT Press, Cambridge, pp. 631-682.

Anderson, J. (1983), *The Architecture of Cognition*, Harvard University Press, Cambridge, MA.

Anscombe, G. E. M. (1957), *Intention*, Blackwell, Oxford.

— (1965), *The Intentionality of Sensation: A Grammatical Feature*, in *Analytical Philosophy* (2nd Series), a cura di R. J. Butler, Blackwell, Oxford, p. 160.

Anton, G. (1899), *Ueber die Selbstwahrnehmung der Herderkrankungen des Gehirs durch den Kranken bei Rindenblindenheit under Rindentaubheit*, «Archiv für Psychiatrie und Nervenkrankheitene», 32, pp. 86-127.

Arnauld, A. (1641), *Quarte Obbiezioni* in Cartesio (1641) *Opere filosofiche*, Vol. II, Biblioteca Universale Laterza, Bari 1986.

Baars, B. (1988), *A Cognitive Theory of Consciousness*, Cambridge University Press, Cambridge.

Bach-y-Rita, P. (1972), *Brain Mechanisms in Sensory Substitution*, Academic Press, New York e Londra.

Ballard, D. e Feldman, J. (1982), *Connectionist Models and Their Properties*, «Cognitive Science», 6, pp. 205-254.

Bechtel, W. e Abrahamsen, A. (1991), *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*, Blackwell, Oxford.

Bennett, J. (1965), *Substance, Reality and Primary Qualities*, «American Philosophical Quarterly», 2, pp. 1-17.
— (1976), *Linguistic Behavior*, Cambridge University Press, Cambridge.
Bentham, J. (1789), *Introduction to Principles of Morals and Legislation*, Londra.
Bick, P.A. e Kinsbourne, M. (1987), *Auditory Hallucinations and Subvocal Speech in Schizophrenic Patients*, «American Journal of Psychiatry», 144, pp. 222-225.
Bieri, P. (1990), *Commentary at the conference: The Phenomenal Mind – How Is It Possible and Why Is It Necessary?*, Zentrum für Interdisziplinäre Forschung, Bielefeld, Germania, 14-17 Maggio.
Birnbaum, L. e Collins, G. (1984), *Opportunistic Planning and Freudian Slips*, «Proceedings, Cognitive Science Society», Boulder, Colorado, pp.124-127.
Bisiach, E. (1988), *The (Haunted) Brain and Consciousness*, in Marcel e Bisiach (1988).
Bisiach, E. e Luzzatti, C. (1978), *Unilateral Neglect of Representational Space*, «Cortex», 14, pp. 129-133.
Bisiach, E. e Vallar, G. (1988), *Hemineglect in Humans*, in F. Boller e J. Grafman (a cura di), *Handbook of Neuropsychology*, Vol. 1, Elsevier, New York.
Bisiach, E., Vallar, G., Perani, D., Papagno, C. e Berti, A. (1986), *Unawareness of Disease Following Lesions of the Right Hemisphere: Anosognosia for Hemiplegia and Anosognosia for Hemianopia*, «Neuropsychologia», 24, pp. 471-482.
Blakemore, C. (1976), *Mechanics of the Mind*, Cambridge University Press, Cambridge (trad. it. di L. Mecacci, *I meccanismi della mente*, Editori Riuniti, Roma 1981).
Block, N. (1978), *Troubles with Functionalism*, in *Perception and Cognition: Issues in the Foundations of Psychology*, a cura di W. Savage, Minnesota Studies in the Philosophy of Science, Vol. IX, pp. 261-326.
— (1981), *Psychologism and Behaviorism*, «Philosophical Review», 90, pp.5-43.
— (1990), *Inverted Earth* in *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind, 1990*, a cura di J. E. Tomberlin, Ridgeview Publishing, Atascadero, California, pp. 53-79.
Boghossian, P. A. e Velleman, J. D. (1989), *Colour as a Secondary Quality*, «Mind», 98, pp. 81-103.
— (1991), *Physicalist Theories of Color*, «Philosophical Review», 100, pp. 67-106.
Booth, W. (1988), *Voodoo Science*, «Science», 240, pp. 274-277.
Borges, J. L. (1944), *Ficciones* (trad. it. di F. Lucentini, *Finzioni*, Einaudi, Torino 1985).
Borgia, G. (1986), *Sexual Selection in Bowerbirds*, «Scientific American», 254, pp. 92-100.
Braitenberg, V. (1984), *Vehicles: Experiments in Synthetic Psychology*,

MIT Press/A Bradford Book, Cambridge (ed. it. a cura di N. Bruno e L. Martinuzzi, *I veicoli pensanti*, Garzanti, Milano 1984).
Breitmeyer, B. G. (1984), *Visual Masking*, Oxford University Press, Oxford.
Broad, C. D. (1925), *Mind and Its Place in Nature*, Routledge & Kegan Paul, Londra.
Brooks, B. A., Yates, J. T. e Coleman, R. D. (1980), *Perception of Images Moving at Saccadic Velocities During Saccades and During Fixation*, «Experimental Brain Research», 40, pp. 71-78.
Byrne, R. e Whiten, A. (1988), *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*, Claredon, Oxford.
Calvanio, R., Petrone, P. N. e Levine, D. N. (1987), *Left visual spatial neglect is both environment-centered and body-centered*, «Neurology», 37, pp. 1179-1183.
Calvin, W. (1983), *The Throwing Madonna: Essays on the Brain*, McGraw-Hill, New York.
— (1986), *The River that Flows Uphill: A Journey from the Big Bang to the Big Brain*, Sierra Club Books, San Francisco.
— (1987), *The Brain as a Darwin Machine*, «Nature», 330, pp. 33-34.
— (1989a), *The Cerebral Symphony: Seashore Reflections on the Structure of Consciousness*, Bantam, New York.
— (1989b), *A Global Brain Theory*, «Science», 240, pp. 1802-1803.
Campion, J., Latto, R. e Smith, Y. M. (1983), *Is Blindsight an Effect of Scattered Light, Spared Cortex, and Near-Threshold Vision?*, «Behavioral and Brain Sciences», 6, pp. 423-486.
Camus, A. (1942), *Le Mythe de Sisyphe*, Gallimard, Parigi (trad. it. di A. Borelli, *Il Mito di Sisifo*, Bompiani, Milano 1984).
Carruthers, P. (1989), *Brute Experience*, «Journal of Philosophy», 86, pp. 258-269.
Castaneda, C. (1968), *The Teachings of Don Juan: A Yaqui Way of Knowledge*, University of California Press, Berkeley (ed. it., *A scuola dallo stregone. Una via Yaqui alla conoscenza*, Astrolabio, Roma).
Castaneda, H.-N. (1967), *Indicators and Quasi-Indicators*, «American Philosophy Quarterly», 4, pp. 85-100.
— (1968), *On the Logic af Attributions of Self-Knowledge to Others*, «Journal of Philosophy», 65, pp. 439-456.
Changeux, J.-P. e Danchin, A. (1976), *Selective Stabilization of Developing Synapses as a Mechanism for the Specifications of Neuronal Networks*, «Nature», 264, pp. 705-712.
Changeux, J.-P. e Dehaene, S. (1989), *Neuronal Models of Cognitive Functions*, «Cognition», 33, pp. 63-109.
Cheney, D. L. e Seyfarth, R. M. (1990), *How Monkeys See the World*, University of Chicago Press, Chicago.
Cherniak, C. (1986), *Minimal Rationality*, MIT Press/A Bradford Book, Cambridge, MA.

Churchland, P. M. (1985), *Reduction, Qualia and the Direct Inspection of Brain States*, «Journal of Philosophy», 82, pp. 8-28.
— (1990), *Knowing Qualia: A Replay to Jackson*, in P. M. Churchland, *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, MIT Press/A Bradford Book, Cambridge, MA, pp. 67-76.
Churchland, P. S. (1981a), *On the Alleged Backwards Referral of Experiences and Its Relevance to the Mind-Body Problem*, «Philosophy of Science», 48, pp. 165-181.
— (1981b), *The Timing of Sensations: Reply to Libet*, «Philosophy of Science», 48, pp. 492-497.
— (1986), *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. MIT Press/A Bradford Book, Cambridge, MA.
Clark, R. W. (1975), *The Life of Bertrand Russell*, Weidenfeld and Nicolson, Londra.
Cohen, L. D., Kipnis, D., Kunkle, E. C. e Kubzansky, P. E. (1955), *Case Report: Observation of a Person with Congenital Insensitivity to Pain*, «Journal of Abnormal and Social Psychology», 51, pp. 333-338.
Cole, D. (1990), *Functionalism and Inverted Spectra*, «Synthese», 82, pp. 207-222.
Crane, H. e Piantanida, T.P. (1983), *On Seeing Reddish Green and Yellowish Blue*, «Science», 222, pp. 1078-1080.
Crick, F. (1984), *Function of the Thalamic Reticular Complex: The Searchlight Hypothesis*, «Proceedings of the National Academy of Sciences», 81, pp. 4586-4590.
Crick, F. e Koch, C. (1990), *Towards a Neurobiological Theory of Consciousness*, «Seminars in the Neurosciences», 2, pp. 263-275.
Damasio, A. R., Damasio, H. e Van Hoesen, G. W. (1982), *Prosopagnosia: Anatomic Basis and Behavioral Mechanisms*, «Neurology», 32, pp. 331-341.
Darwin, C. (1871), *The Descent of Man, and Selection in Relation to Sex*, 2 vol., Murray, Londra (trad. it. di M. Lessona, *L'origine dell'uomo e la scelta sessuale*, Rizzoli, Milano 1982).
Davis, W. (1985), *The Serpent and the Rainbow*, Simon & Schuster, New York.
— (1988a), *Passage of Darkness: The Ethnobiology of the Haitian Zombie*, University of North Carolina Press, Chapel Hill e Londra.
— (1988b), *Zombification*, «Science», 240, pp. 1715-1716.
Dawkins, M. S. (1980), *Animal Suffering: The Science of Animal Welfare*, Chapman & Hall, Londra.
— (1987), *Minding and Mattering* in *Mindwaves*, a cura di C. Blakemore e S. Greenfield, Blackwell, Oxford, pp. 150-160.
— (1990), *From an Animal's Point of View: Motivation, Fitness, and Animal Welfare*, «Behavioral and Brain Sciences», 13, pp. 1-61.
Dawkins, R. (1976), *The Selfish Gene*, Oxford University Press, Oxford (trad. it. di G. Corte e A. Serra, *Il gene egoista*, Mondadori, Milano 1992).
— (1982), *The Extended Phenotype*, Freeman, San Francisco (trad. it. di

R. Marpurgo, *Il fenotipo esteso. Il gene come unità di selezione*, Zanichelli, Bologna 1986).
— (1986), *The Blind Watchmaker*, Norton, New York (trad. it. di L. Sosio, *L'orologiaio cieco*, Rizzoli Milano 1988).
de Sousa, R. (1976), *Rational Homunculi*, in *The Identity of Persons*, a cura di Amelie O. Rorty, University of California Press, Berkeley, pp. 217-238.
Dennett, D. C. (1969), *Content and Consciousness*, Routledge & Kegan Paul, Londra (trad. it. di G. Pacini Mugnai, *Contenuto e coscienza*, Il Mulino, Bologna 1992).
— (1971), *Intentional Systems*, «Journal of Philosophy», 8, pp. 87-106 (ristampato in Dennett [1978a], cap. 1, trad. it., «I sistemi intenzionali»).
— (1974), *Why the Law of Effect Will Not Go Away*, «Journal of the Theory of Social Behaviour», 5, pp. 169-187 (ristampato in Dennett [1978a], cap. 5, trad. it., «Perché la legge dell'effetto non sarà abbandonata»).
— (1975), *Are Dreams Experiences?*, «Philosophical Review», LXXIII, pp. 151-171 (ristampato in Dennett [1978a], cap. 8, trad. it. «I sogni sono esperienze?»).
— (1978a), *Brainstorms*, Bradford Books, Montgomery (trad. it. di L. Colasanti, *Brainstorms*, Adelphi, Milano 1991).
— (1978b), «Skinner Skinned», cap. 4 in Dennett (1978a), trad. it. «Skinner scorticato».
— (1978c), «Two Approaches to Mental Images», cap. 10 in Dennett (1978a), trad. it. «Due approcci alle immagini mentali».
— (1978d), «Where Am I?», cap. 17 in Dennett (1978a), trad. it. «Dove sono?».
— (1979a), *On the Absence of Phenomenology* in *Body, Mind and Method: Essays in Honor of Virgil Aldrich*, a cura di D. Gustafson e B. Tapscott, Reidel, Dordrecht, 1979.
— (1979b), Recensione a Popper e Eccles, *The Self and Its Brain: An Argument for Interactionism*, in «Journal of Philosophy», 76, pp. 91-97.
— (1981a), «Riflessioni» su «Software» in Hofstadter e Dennett, 1981.
— (1981b), *Wondering Where the Yellow Went* (commento alle Carus Lectures di W. Sellars), «Monist», 64, pp. 159-108.
— (1982a), *How to Study Human Consciousness Empirically, or Nothing Comes to Mind*, «Synthese», 59, pp. 159-180.
— (1982b), *Why We Think What We Do about Why We Think What We Do: Discussion on Goodman's "On Thoughts without Words"*, «Cognition», 12, pp. 219-227.
— (1982c), *Comments on Rorty*, «Synthese», 59, pp. 349-356.
— (1982d), *Notes on Prosthetic Imagination*, «New Boston Review», giugno, pp. 3-7.
— (1983), *Intentional Systems in Cognitive Ethology: The "Panglossian*

Paradigm" Defended, «Behavioral and Brain Sciences», 6, pp. 343-390.
— (1984a), *Elbow Room: The Varieties of Free Will Worth Wanting*, MIT Press/A Bradford Book, Cambridge, MA.
— (1984b), *Carving the Mind at Its Joints*, recensione a Fodor, *The Modularity of Mind*, in «Contemporary Phychology», 29, pp. 285-286.
— (1985a), *Can Machines Think?* in «How We Know», a cura di M. Shafto, Harper & Row, New York, pp. 121-145.
— (1985b), *Music of the Hemispheres*, recensione a M. Gazzaniga, *The Social Brain*, in «New York Times Book Review», 17.11.1985, p. 53.
— (1986), *Julian Jaynes' Software Archeology*, «Canadian Psychology», 27, pp. 149-154.
— (1987a), *The Intentional Stance*, MIT Press/A Bradford Book, Cambridge, MA.
— (1987b), *The Logical Geography of Computational Approaches: A View from the East Pole*, in *Problems in the Representation of Knowledge*, a cura di M. Harnish e M. Brand, University of Arizona Press, Tucson.
— (1988a) *Quining Qualia*, in Marcel e Bisiach (a cura di) (1988), pp. 42-77.
— (1988b), *When Philosophers Encounter AI*, «Daedalus», 117, pp. 283-296; ristampato in Graubard (1988).
— (1988c), *Out of the Armchair and Into the Field*, «Poetics Today», 9, numero speciale su «Intepretation in Context in Science and Culture», pp. 205-222.
— (1988d), *The Intentional Stance in Theory and Practice*, in Whiten e Byrne (1988), pp. 180-202.
— (1988e), *Science, Philosophy and Interpretation*, «Behavioral and Brain Sciences», 11, pp. 535-546.
— (1988f), *Why Everyone Is a Novelist*, «Times Literary Supplement», 16-22 settembre.
— (1989a), *Why Creative Intelligence Is Hard to Find*, commenti a Whiten e Byrne, «Behavioral and Brain Sciences», 11, p. 253.
— (1989b), *The Origins of Selves*, «Cogito», 2, pp. 163-173.
— (1989c), *Murmurs in the Cathedral*, recensione a R. Penrose *The Emperor's New Mind*, in «Times Literary Supplement», 29 settembre - 5 ottobre, pp. 1066-1068.
— (1989d), *Cognitive Ethology: Hunting for Bargains or a Wild Goose Chase?* in *Goals, Own Goals and No Goals: A Debate on Goal-Directed And Intentional Behaviour*, a cura di A. Montefiore e D. Noble, Unwin Hyman, Londra.
— (1990a), *Memes and the Exploitation of Imagination*, «Journal of Aesthetics and Art Criticism», 48, pp. 127-135.
— (1990b), *Thinking with a Computer*, in *Image and Understanding*, a cura di H. Barlow, Cambridge University Press, Cambridge, pp. 297-309.

— (1990c), *Betting Your Life on an Algorithm*, commenti a Penrose, «Behavioral and Brain Sciences», 13, p. 660.
— (1990d), *The Interpretation of Texts, People, and Other Artifacts*, «Philosophy and Phenomenological Research», 50, pp. 177-194.
— (1990e), *Two Black Boxes: A Fable*, Tufts University Center for Cognitive Studies Preprint, Novembre.
— (1991a), *Real Patterns*, «Journal of Philosophy», 89, pp. 27-51.
— (1991b), *Producing Future by Telling Stories*, in *Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*, a cura di K. M. Ford e Z. Pylyshyn, Ablex Series in Theoretical Issues in Cognitive Science, Ablex, Norwood.
— (1991c), *Mother Nature versus the Walking Encyclopedia* in *Philosophy and Connectionist Theory*, a cura di W. Ramsey, S. Stich e D. Rumelhart, Erlbaum, Hillsdale.
— (1991d), *Two Contrasts: Folk Craft versus Folk Science and Belief versus Opinion*, in *The Future of Folk Psychology: Intentionality and Cognitive Science*, a cura di J. Greenwood, Cambridge University Press, Cambridge, 1991.
— (1991e), *Granny's Campaign for Safe Science* in *Fodor and His Critics*, a cura di G. Rey e B. Loewer, Blackwell, Oxford.
Dennett, D. e Kinsbourne, M. (in stampa), *Time and the Observer: The Where and When of Consciousness in the Brain*, «Behavioral and Brain Sciences».
Descartes, R. (1637), *Discours de la méthode*, Leida (trad. it. di A. Carlini, *Discorso sul metodo*, Piccola biblioteca filosofica, Laterza, Bari 1938).
— (1641), *Meditationes de prima philosophia*, Parigi (trad. it. di A. Tilgher, in Cartesio *Opere filosofiche*, vol. 2, Laterza, Bari 1967).
— (1664), *L'Homme*, Girard, Parigi (trad. it. *L'uomo* di E. Garin, in Cartesio *Opere filosofiche*, vol. 1, Laterza, Bari 1967).
— (1970), *Philosophical Letters*, a cura di A. Kenny, Clarendon Press, Oxford.
Dreyfus, H. L. (1979), *What Computers Can't Do* (2ª edizione), Harper & Row, New York (trad. it. di G. Alessandrini, *Che cosa non possono fare i computer. I limiti dell'intelligenza artificiale*, Armando, Roma).
Dreyfus, H. L. e Dreyfus, S. E. (1988), *Making a Mind Versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint*, in Graubard (1988).
Eccles, J. C. (1985), *Mental Summation: The Timing of Voluntary Intentions by Cortical Activity*, «Behavioral and Brain Sciences», 8, pp. 542-547.
Eco, U. (1990), *After Secret Knowledge*, «Times Literary Supplement», 22-28 giugno, p. 666; *Some Paranoid Readings*, «Times Literary Supplement», 29 giugno - 5 luglio, p. 694.
Edelman, G. (1987), *Neural Darwinism*, Basic Books, New York.
— (1989), *The Remembered Present: A Biological Theory of Consciousness*,

Basic Books, New York (trad. it. di L. Sosio, *Il presente ricordato*, Rizzoli, Milano 1991).
Efron, R. (1967), *The Duration of the Present*, «Proceedings of the New York Academy of Science», 8, pp. 542-543.
Eldredge, N. e Gould, S. J. (1972), *Punctuated Equilibria: An Alternative to Phyletic Gradualism*, in *Models in Paleobiology*, a cura di T. J. M. Schopf, Freeman Cooper, San Francisco, pp. 82-115.
Ericsson, K. A. e Simon, H. A. (1984), *Protocol Analysis: Verbal Reports as Data*, MIT Press/A Bradford Book, Cambridge, MA.
Evans, G. (1982), in *The Varieties of Reference*, a cura di J. McDowell, Oxford University Press, Oxford.
Ewert, J.-P. (1987), *The Neuroethology of Releasing Mechanisms: Prey-catching in Toads*, «Behavioral and Brain Sciences», 10, pp. 337-405.
Farah, M. J. (1988), *Is Visual Imagery Really Visual? Overlooked Evidence from Neuropsychology*, «Psychological Review», 95, pp. 307-317.
Farrell, B. A. (1950), *Experience*, Mind, 59, pp. 170-198.
Fehling, M., Baars, B. e Fisher, C. (1990), *A Functional Role of Repression in an Autonomous, Resource-constrained Agent* in «Proceedings of Twelfth Annual Conference of the Cognitive Science Society», Erlbaum, Hillsdale.
Fehrer, E. e Raab, D. (1962), *Reaction Time to Stimuli Masked by Metacontrast*, «Journal of Experimental Psychology», 63, pp. 143-147.
Feynmann, R. (1985), *Surely You're Joking, Mr. Feynmann!*, Norton, New York (trad. it. di S. Coyaud, *Sta scherzando, Mr. Feynmann!*, Zanichelli, Bologna 1988).
Finke, R. A., Pinker, S. e Farah, M. J. (1989), *Reinterpreting Visual Patterns in Mental Imagery*, «Cognitive Science», 13, pp. 51-78.
Flanagan, O. (1991), *The Science of the Mind* (2ª edizione), MIT Press/A Bradford Book, Cambridge, MA.
Flohr, H. (1990), *Brain Processes and Phenomenal Consciousness: A New and Specific Hypothesis*, presentato alla conferenza «The Phenomenal Mind – How Is It Possible and Why Is It Necessary?», Zentrum für Interdisziplinäre Forschung, Bielefeld, Germania, 14-17 maggio.
Fodor, J. (1975), *The Language of Thought*, Crowell, Scranton.
— (1983), *The Modularity of Mind*, MIT Press/A Bradford Book, Cambridge, MA (trad. it. di R. Luccio, *La mente modulare. Saggio di psicologia delle facoltà*, Il Mulino, Bologna 1988).
— (1990), *A Theory of Content, and Other Essays*, MIT Press/A Bradford Book, Cambridge, MA.
Fodor, J. e Pylyshyn, Z. (1988), *Connectionism and Cognitive Architecture: A Critical Analysis*, «Cognition», 28, pp. 3-7.
Fox, I. (1989), *On the Nature and Cognitive Function of Phenomenal Content - Part One*, «Philosophical Topics», 17, pp. 81-117.
French, R. (1991), *Subcognition and the Turing Test*, «Mind», in stampa.

Freud, S. (1922), *L'Io e l'Es* (ed. Boringhieri, Torino).
Freyd, J. (1989), *Dynamic Mental Representations*, «Psychological Review», 94, pp. 427-438.
Fuster, J. M. (1981), *Prefrontal Cortex in Motor Control*, in *Handbook of Physiology, Section 1: The Nervous System, Vol. II: Motor Control*, American Physiological Society, pp. 1149-1178.
Gardner, H. (1975), *The Shattered Mind*, Knopf, New York.
Gardner, M. (1981), *The Laffer Curve and Other Laughs in Current Economics*, «Scientific American», 245, dicembre, pp. 18-31, ristampato in Gardner (1986).
— (1986), *Knotted Doughnuts and Other Mathematical Diversions*, W. H. Freeman, San Francisco.
Gazzaniga, M. (1978), *Is Seeing Believing: Notes on Clinical Recovery*, in *Recovery From Brain Damage: Research and Theory*, a cura di S. Finger, Plenum Press, New York, pp. 409-414.
— (1985), *The Social Brain: Discovering the Networks of the Mind*, Basic Books, New York (trad. it. *Il cervello sociale*, Giunti-Barbera, Firenze, 1989).
Gazzaniga, M. e Ledoux, J. (1978), *The Integrated Mind*, Plenum Press, New York.
Geldard, F. A. (1977), *Cutaneous Stimulis, Vibratory and Saltatory*, «Journal of Investigative Dermatology», 69, pp. 83-87.
Geldard, F. A. e Sherrick, C. E. (1972), *The Cutaneous "Rabbit": A Perceptual Illusion*, «Science», 178, pp. 178-179.
— (1983), *The Cutaneous Saltatory Area and Its Presumed Neural Base*, «Perception and Psychophysics», 33, pp. 299-304.
— (1986), *Space, Time and Touch*, «Scientific American», 254, pp. 90-95.
Gert, B. (1965), *Imagination and Verifiability*, «Philosophical Studies», 16, pp. 44-47.
Geshwind, N. e Fusillo, M. (1966), *Color-naming Defects in Association with Alexia*, «Archives of Neurology», 15, pp. 137-146.
Gide, A. (1948), *Les Faux Monnayeurs*, Gallimard, Paris (trad. it. di O. Del Buono, *I Falsari*, Bompiani, Milano 1978).
Goodman, N. (1978), *Ways of Worldmaking*, Harvester, Hassocks, Sussex (trad. it. di C. Marletti, *Vedere e costruire il mondo*, Laterza, Bari 1988).
Goody, J. (1977), *The Domestication of the Savage Mind*, Cambridge University Press, Cambridge (ed. it. *L'addomesticamento del pensiero selvaggio*, Franco Angeli, Milano 1990).
Gould, S. (1980), *The Panda's Thumb*, Norton, New York (trad. it. di S. Cabib, *Il pollice del Panda. Riflessioni sulla storia naturale*, Editori Riuniti, Roma).
Gouras, P. (1984), *Color Vision*, in «Progress in Retinal Research», Vol. 3, a cura di N. Osborn e J. Chader, Pergamon Press, London, pp. 227-261.
Graubard, S. R. (1988). *The Artificial Intelligence Debate: False Starts,*

Real Foundations (ristampa di «Daedalus», 117, inverno 1988), MIT Press, Cambridge, MA.
Grey Walter, W. (1963), *Presentation to the Osler Society*, Oxford University.
Grice, H. P. (1957), *Meaning*, «Philosophical Review», 66, pp. 377-388.
— (1969), *Utterer's Meaning and Intentions*, «Philosophical Review», 78, pp. 147-177.
Hacking, I. (1990), *Signing*, recensione a Sacks (1989), «London Review of Books», 5 aprile, pp. 3-6.
Hampl, P. (1989), *The Lax Habits of the Free Imagination*, «New York Times Book Review», 5 marzo, pp. 1, 37-39, brani tratti da *The Houghton Mifflin Anthology of Short Fiction*, a cura di Hampl, 1989, Hougthon Mifflin, Boston.
Handford, M. (1987), *Where's Waldo?*, Little, Brown, Boston (trad. it. di G. Mantegazza, *Ubaldo, dove sei?*, Piccoli, Milano 1987).
Hardin, C. L. (1988), *Color for Philosophers: Unweaving the Rainbow*, Hackett, Indianapolis.
— (1990), *Color and Illusion*, presentato alla conferenza «The Phenomenal Mind – How Is It Possible and Why Is It Necessary?», Zentrum für Interdisziplinäre Forschung, Bielefeld, Germania, 14-17 maggio.
Harman, G. (1990), *The Intrinsic Quality of Experience*, in *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, a cura di J. E. Tomberlin, Ridgeview, Atascadero, pp. 31-52.
Harnad, S. (1982), *Consciousness: An Afterthought*, «Cognition and Brain Theory», 5, pp. 29-47.
— (1989), *Editorial Commentary*, «Behavioral and Brain Sciences», 12, p. 183.
Haugeland, J. (1981), *Mind Design: Philosophy, Psychology, Artificial Intelligence*, Bradford Books, Montgomery (ed. it. *Progettare la mente*, Il Mulino, Bologna 1989).
— (1985), *Artificial Intelligence: The Very Idea*, MIT Press/A Bradford Book, Cambridge, MA. (trad. it. di V. Sala, *Intelligenza Artificiale*, Bollati Boringhieri, Torino 1988).
Hawking, S. (1988), *A Brief History of Time*, Bantam, New York (trad. it. di L. Sosio, *Dal big bang ai buchi neri. Breve storia del tempo*, Rizzoli, Milano 1988).
Hayes, P. (1979), *The Naive Physics Manifesto*, in *Expert Systems in the Microelectronic Age*, a cura di D. Michie, Edinburgh University Press, Edinburgh.
Hayes-Roth, B. (1985), *A Blackboard Architecture for Control*, «Artificial Intelligence», 26, pp. 251-321.
Hebb, D. (1949), *The Organization of Behavior: A Neuropsychological Theory*, Wiley, New York (trad. it. di R. Grande, *L'organizzazione del comportamento. Una teoria neuropsicologica*, F. Angeli, Milano 1975).
Hilbert, D. R. (1987), *Color and Color Perception: A Study in Anthropocen-*

tric Realism, Stanford University; Center for the Study of Language and Information.
Hintikka, J. (1962), *Knowledge and Belief.* Cornell University Press, Ithaca.
Hinton, G. E. e Nowland, S. J. (1987), *How Learning Can Guide Evolution, Complex Systems, I,* Technical Report CMU-CS-86-128, Carnegie Mellon University, pp. 495-502.
Hobbes, T. (1651), *Leviathan,* Parigi (ed. it. a cura di A. Pacchi, *Leviatano,* Laterza, Bari 1989).
Hoffman, R. E. (1986), *What Can Schizophrenic "Voices" Tell Us?,* «Behavioral and Brain Sciences», pp. 535-548.
Hoffman, R. E. e Kravitz, R. E. (1987), *Feedforward Action Regulation and the Experience of Will,* «Behavioral and Brain Sciences», 10, pp. 782-783.
Hofstadter, D. R. (1981a), *The Turing Test: A Coffehouse Conversation,* in *Metamagical Themas,* «Scientific American», maggio, ristampato in Hofstadter e Dennett (1981), pp. 69-92 (trad. it. pp. 76-97).
— (1981b), *Riflessioni [su Nagel]* in Hofstadter e Dennett (1981), pp. 403-414 (trad. it. pp. 392-403).
— (1983), *The Architecture of Jumbo,* «Proceedings of the Second Machine Learning Workshop, Monticello.
— (1985), *On the Seeming Paradox of Mechanizing Creativity,* in *Metamagical Themas,* Basic Books, New York, pp. 526-546 (trad. it. *Si può meccanizzare la creatività?* in «Le Scienze», novembre 1982).
Hofstadter, D. R. e Dennett, D. C. (1981), *The Mind's I: Fantasies and Reflections on Self and Soul,* Basic Books, New York, pp. 191-201 (trad. it. di G. Longo, *L'io della mente,* Adelphi, Milano 1985, pp. 189-198).
Holland, J. H. (1975), *Adaptation in Natural and Artificial Systems,* University of Michigan Press, Ann Arbor.
Holland, J. H., Holyoak, K. J., Nisbett, R. E. e Thagard, P. R. (1986), *Induction: Processes of Inference, Learning, and Discovery,* MIT Press/A Bradford Book, Cambridge, MA.
Honderich, T. (1984), *The Time of a Conscious Sensory Experience and Mind-Brain Theories,* «Journal of Theoretical Biology», 110, pp. 115-129.
Howell, R. (1979), *Fictional Objects: How They Are and How They Aren't,* in *Body, Mind and Method,* a cura di D. F. Gustafson e B. L. Tapscott, D. Reidel, Dordrecht, pp. 241-294.
Hughlings Jackson, J. (1915), *Hughlings Jackson on Aphasia and Kindred Affections of Speech,* «Brain» 38, pp. 1-190.
Hume, D. (1739), *Treatise on Human Nature,* John Noon, Londra (trad. it. a cura di E. Lecaldano, *Trattato sulla natura umana,* Laterza, Bari 1975).
Humphrey, N. (1972), *"Interest" and "Pleasure": Two Determinants of a Monkey's Visual Preferences,* «Perception», 1, pp. 395-416.
— (1976), *The Colour Currency of Nature,* in *Colour for Architecture,* a cura

di T. Porter e B. Mikellides, Studio-Vista, Londra, pp. 147-161, ristampato in Humphrey (1983a).
— (1983a), *Consciousness Regained*, Oxford University Press, Oxford.
— (1983b), *The Adaptiveness of Mentalism?* commento a Dennett (1983), «Behavioral and Brain Sciences», 6, p. 366.
— (1986), *The Inner Eye*, Faber & Faber, Londra (trad. it. di L. Montixi Comoglio, *L'occhio della mente. Ovvero perché gli animali non si guardano allo specchio*, Instar Libri, Torino 1992).
— (di prossima pubblicazione), *A History of the Mind*, Simon & Schuster, New York.
Humphrey, N. e Dennett, D. C. (1989), *Speaking for Our Selves: An Assessment of Multiple Personality Disorder*, «Raritan», 9, pp. 68-98.
Humphrey, N. e Keeble, G. (1978), *Effects of Red Light and Loud Noise on the Rates at Which Monkeys Sample the Sensory Environment*, «Perception», 7, p. 343.
Hundert, E. (1987), *Can Neuroscience Contribute to Philosophy?* in *Mindwaves*, a cura di C. Blakemore e S. Greenfield, Blackwell, Oxford, pp. 407-429 (ristampato come cap. 7 di Hundert, *Philosophy, Psychiatry, and Neuroscience: Three Approaches to the Mind*, Clarendon, Oxford 1989).
Huxley, T. (1874), *On the Hypothesis that Animals Are Automata*, in *Collected Essays*, London, 1893-1894.
Jackendoff, R. (1987), *Consciousness and the Computational Mind*, MIT Press/A Bradford Book, Cambridge, MA (trad. it. di S. Gozzano, *Coscienza e mente computazionale*, Il Mulino, Bologna 1990).
Jackson, F. (1982), *Epiphenomenal Qualia*, «Philosophical Quarterly», 32, pp. 127-136.
Jacob, F. (1982), *The Possible and the Actual*, University of Washington Press, Seattle.
Janlert, L.-E. (1985), *Studies in Knowledge Representation*, Institute of Information Processing, Umea, Sweden.
Jarrell, R. (1963), *The Bat-Poet*, Macmillan, New York.
Jaynes, J. (1976), *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, Houghton Mifflin, Boston (trad. it. di L. Sosio, *Il crollo della mente bicamerale e l'origine della coscienza*, Adelphi, Milano 1984).
Jerison, H. (1973), *Evolution of the Brain and Intelligence*, Academic Press, New York.
Johnson-Laird, P. (1983), *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge University Press, Cambridge (trad. it. di A. Mazzocco, *Modelli mentali*, Il Mulino, Bologna 1988).
— (1988), *A Computational Analysis of Consciousness*, in Marcel e Bisiach (1988).

Julesz, B. (1971), *Foundations of Cyclopean Perception*, University of Chicago Press, Chicago.
Keller, H. (1908), *The World I Live In*, Century Co, New York (trad. it. *Il mondo in cui vivo*, Bocca, Milano 1944).
Kinsbourne, M. (1974), *Lateral Interactions in the Brain*, in *Hemisphere Disconnection and Cerebral Function*, a cura di M. Kinsbourne e W. L. Smith, Charles C. Thomas, Springfield, pp. 239-259.
— (1980), *Brain-based Limitations on Mind*, in *Body and Mind: Past, Present and Future*, a cura di R. W. Rieber, Academic Press, New York, pp. 155-175.
Kinsbourne, M. e Hicks, R. E. (1978), *Functional Cerebral Space: A Model for Overflow, Transfer and Interference Effects in Human Performance: A Tutorial Review*, in *Attention and Performance*, a cura di J. Requin, 7, Erlbaum, Hillsdale, pp. 345-362.
Kinsbourne, M. e Warrington, E. K. (1963), *Jargon Aphasia*, «Neuropsychologia», 1, pp. 27-37.
Kirman, B. H. e altri (1968), *Congenital Insensitivity to Pain in an Imbecile Boy*, «Developmental Medicine and Child Neurology», 10, pp. 57-63.
Kitcher, P. (1979), *Phenomenal Qualities*, «American Philosophical Quarterly», 16, pp. 123-129.
Koestler, A. (1967), *The Ghost in the Machine*, Macmillan, New York (trad. it. *Il fantasma dentro la macchina*, SEI, Torino 1971).
Kohler, I. (1961), *Experiments with Goggles*, «Scientific American», 206, pp. 62-86.
Kolers, P. A. (1972), *Aspects of Motion Perception*, Pergamon Press, London.
Kolers, P. A. e von Grünau, M. (1976), *Shape and Color in Apparent Motion*, «Vision Research», 16, pp. 329-335.
Kosslyn, S. M. (1980), *Image and Mind*, Harvard University Press, Cambridge, MA.
Kosslyn, S. M., Holtzman, J. D., Gazzaniga, M. S. e Farah, M. J. (1985), *A Computational Analysis of Mental Imagery Generation: Evidence for Functional Dissociation in Split Brain Patients*, «Journal of Experimental Psychology: General», 114, pp. 311-341.
Lackner, J. R. (1988), *Some Proprioceptive Influences on the Perceptual Representation of Body Shape and Orientation*, «Brain», 111, pp. 281-297.
Langton, C. G. (1989), *Artificial Life*, Addison-Wesley, Redwood City.
Larkin, S. e Simon, H. A. (1987), *Why a Diagram Is (Sometimes) Worth Ten Thousand Words*, «Cognitive Science», 11, pp. 65-100.
Leiber, J. (1988), *"Cartesian" Linguistics?*, «Philosophia», 118, pp. 309-346.
— (1991), *Invitation to Cognitive Science*, Blackwell, Oxford.
Leibniz, G. W. (1714), *Monadologie*, pubblicata sotto il titolo *Principia philosophiae seu theses in gratiam Principis Eugenii conscriptae*, «Acta Eruditorum», 1721; trad. it. in *Monadologia e Discorso di metafisica*, Laterza, Roma-Bari 1986.

Levelt, W. (1989), *Speaking*, MIT Press/A Bradford Book, Cambridge, MA.
Levy, J. e Trevarthen, C. (1976), *Metacontrol of Hemispheric Function in Human Split-Brain Patients*, «Journal of Experimental Psychology: Human Perception and Performance», 3, pp. 299-311.
Lewis, D. (1978), *Truth in Fiction*, «American Philosophical Quarterly», 15, pp. 37-46.
— (1979), *Attitudes De Dicto and De Se*, «Philosophical Review», 78, pp. 513-543.
— (1988), *What Experience Teaches, Proceedings of the Russellian Society of the University of Sidney*, ristampato in *Mind and Cognition: A Reader*, a cura di W. Lycan, Backwell, Oxford 1990.
Liberman, A. e Studdert-Kennedy, M. (1977), *Phonetic Perception*, in *Handbook of Sensory Physiology*, Vol.8, *Perception*, a cura di R. Held, H. Leibowitz e H.-L. Teuber, Spinger-Verlag, Heidelberg.
Libet, B. (1965), *Cortical Activation in Conscious and Unconscious Experience*, «Perspectives in Biology and Medicine», 9, pp. 77-86.
— (1981), *The Experimental Evidence for Subjective Referral of a Sensory Experience backward in Time: Reply to P. S. Churchland*, «Philosophy of Science», 48, pp. 182-197.
— (1982), *Brain Stimulation in the Study of Neuronal Functions for Conscious Sensory Experiences*, «Human Neurobiology», 1, pp. 235-242.
— (1985a), *Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action*, «Behavioral and Brain Sciences», 8, pp. 529-566.
— (1985b), *Subjective Antedating of a Sensory Experience and Mind-Brain Theories*, «Journal of Theoretical Biology», 114, pp. 563-570.
— (1987), *Are the Mental Experiences of Will and Self-control Significant for the Performance of a Voluntary Act?*, «Behavioral and Brain Sciences», 10, pp. 783-786.
— (1989), *The Timing of a Subjective Experience*, «Behavioral and Brain Sciences», 12, pp. 183-185.
Libet, B., Wright, E. W., Feinstein, B. e Pearl, D. K. (1979), *Subjective Referral of the Timing for a Conscious Sensory Experience*, «Brain», 102, pp. 193-224.
Liebmann, S. (1927), *Ueber das Verhalten fahrbiger Formen bei Heligkeitsgleichtheit von Figur und Grund*, «Psychologie Forschung», 9, pp. 200-253.
Livingstone, M. S. e Hubel, D. H. (1987), *Psychophysical Evidence for Separate Channels for the Perception of Form, Color, Movement, and Depth*, «Journal of Neuroscience», 7, pp. 346-368.
Lloyd, M. e Dybas, H. S. (1966), *The Periodical Cicada Problem*, «Evolution», 20, pp. 132-149.
Loar, B. (1990), *Phenomenal Properties*, in «Philosophical Perspectives, 4: Action Theory and Philosophy of Mind», a cura di J. E. Tomberlin, Ridgeview, Atascadero, pp. 81-108.
Locke, J. (1690), *Essay Concerning Human Understanding*, Basset, Londra (trad. it. *Saggi sull'intelletto umano*, Laterza, Roma-Bari 1990).

Lockwood, M. (1989), *Mind, Brain and the Quantum*, Blackwell, Oxford.
Lodge, D. (1988), *Nice Work*, Secker and Warburg, Londra (trad. it. di M. Buckwell e R. Palazzi, *Ottimo lavoro, professore!*, Bompiani, Milano 1991).
Lycan, W. (1973), *Inverted Spectrum*, «Ratio», 15, pp. 315-319.
— (1990), *What Is the Subjectivity of the Mental?*, in *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, a cura di J. E. Tomberlin, Ridgeview, Atascadero, pp. 109-130.
Marais, E. N. (1937), *The Soul of the White Ant*, Methuen, Londra (trad. it. di Bacchi Wilcock, *L'anima della formica bianca*, Adelphi, Milano 1984).
Marcel, A. J. (1988), Phenomenal Experience and Functionalism, in Marcel e Bisiach (1988), pp. 121-158.
— (in stampa), *Slippage in the Unity of Consciousness*, in *Perception Without Awareness: Cognitive, Clinical and Social Perspectives*, a cura di R. Bornstein e T. Pittman, Guilford Press, New York.
Marcel, A. e Bisiach, E. (1988), (a cura di), *Consciousness in Contemporary Science*, Oxford University Press, New York.
Margolis, H. (1987), *Patterns, Thinking, and Cognition*, University of Chicago Press, Chicago.
Margulis, L. (1970), *The Origin of Eukaryotic Cells*, Yale University Press, New Haven.
Marks, C. (1980), *Commissurotomy, Consciousness and Unity of Mind*, MIT Press/A Bradford Book, Cambridge, MA.
Marler, P. e Sherman, V. (1983), *Song Structure Without Auditory Feedback: Emendations of the Auditory Template Hypothesis*, «Journal of Neuroscience», 3, pp. 517-531.
Marr, D. (1982), *Vision*, Freeman, San Francisco.
Maynard Smith, J. (1978), *The Evolution of Sex*, Cambridge University Press, Cambridge.
— (1989), *Sex, Games, and Evolution*, Harvester, Brighton, Sussex.
McClelland, J. e Rumelhart, D. (a cura di) (1986), *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, 2 voll., MIT Press/A Bradford Book, Cambridge, MA (trad. it. parziale, *La microstruttura dei processi cognitivi. Elaborazione distribuita in parallelo*, Il Mulino, Bologna 1990).
McCulloch, W. S. e Pitts, W. (1943), *A Logical Calculus for the Ideas Immanent in Nervous Activity*, «Bulletin of Mathematical Biophysics», 5, pp.115-133.
McGinn, C. (1989), *Can We Solve the Mind-Body Problem?*, «Mind», 98, pp.349-366.
— (1990), *The Problem of Consciousness*, Blackwell, Oxford.
McGlynn, S. M. e Schacter, D. L. (1989), *Unawareness of Deficits in Neuropsychological Syndromes*, «Journal of Clinical and Experimental Neuropsychology», 11, pp. 143-205.
McGurk, H. e Macdonald, R. (1979), *Hearing Lips and Seeing Voices*, «Nature», 264, pp. 746-748.

McLuhan, M. (1967), *The Medium Is the Message*, Bantam, New York.
Mellor, H. (1981), *Real Time*, Cambridge Univ. Press, Cambridge.
Menzel, E. W., Savage-Rumbaugh, E. S. e Lawson, J. (1985), *Chimpanzee (Pan troglodytes) Spatial Problem Solving with the Use of Mirrors and Televised Equivalents of Mirrors*, «Journal of Comparative Psychology», 99, pp. 211-217.
Millikan, R. (1990), *Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox*, «Philosophical Review», 99, pp. 323-354.
Minsky, M. (1975), *A Framework for Representing Knowledge*, Memo 3306, AI Lab, MIT, Cambridge, MA (estratti pubblicati in Haugeland [1981], pp. 95-128; trad. it. *Un sistema per la rappresentazione della conoscenza*, pp. 107-142).
— (1985), *The Society of Mind*, Simon & Schuster, New York (trad. it. di G. Longo, *La società della mente*, Adelphi, Milano 1989).
Mishkin, M., Ungerleider, L. G. e Macko, K. A. (1983), *Object Vision and Spatial Vision: Two Cortical Pathways*, «Trends in Neuroscience», 64, pp. 370-375.
Monod, J. (1970), *Le Hasard et la nécessité*, Editions du Seuil, Parigi (trad. it. di A. Busi, *Il caso e la necessità*, A. Mondadori, Milano 1970).
Morris, R. K., Rayner, K. e Pollatsek, A. (1990), *Eye Movement Guidance in Reading: The Role of Parafoveal and Space Information*, «Journal of Experimental Psychology: Human Perception and Performance», 16, pp. 268-281.
Mountcastle, V. B. (1978), *An Organizing Principle for Cerebral Function: The Unit Module and the Distributed System*, in *The Mindful Brain*, a cura di G. Edelman e V. B. Mountcastle, MIT Press, Cambridge, MA., pp. 7-50.
Nabokov, V. (1930), *Zaschita Luzhina*, in *Sovremennye Zapiski*, Parigi; messo in forma di libro da Slovo, Berlino, 1930 (edizione inglese, *The Defence*, Popular Library 1964).
Nagel, T. (1971), *Brain Bisection and the Unity of Consciousness*, «Synthese», 22, pp. 396-413 (ristampato in Nagel, *Mortal Questions*, Cambridge University Press, Cambridge, 1979; trad. it. *Questioni mortali*, Il Saggiatore, Milano).
— (1974), *What Is It Like to Be a Bat?*, «Philosophical Review», 83, pp. 435-450 (trad. it. in Hofstadter e Dennett [1981]).
— (1986), *The View from Nowhere*, Oxford University Press, Oxford (trad. it. *Uno sguardo da nessun luogo*, Il Saggiatore, Milano 1988).
Neisser, U. (1967), *Cognitive Psychology*, Appleton-Century-Crofts, New York (trad. it. *Psicologia cognitivista*, Giunti, Firenze 1976).
— (1981), *John Dean's Memory: A Case Study*, «Cognition», 9, pp. 1-22.
— (1988), *Five Kinds of Self-Knowledge*, «Philosophical Psychology», 1, pp. 35-39.
Nemirow, L. (1990), *Physicalism and the Cognitive Role of Acquaintance*, in *Mind and Cognition*, a cura di W. Lycan, Blackwell, Oxford, pp. 490-499.

Neumann, O. (1990), *Some Aspects of Phenomenal Consciousness and Their Possible Functional Correlates*, presentato alla conferenza «The Phenomenal Mind – How Is It Possible and Why Is It Necessary?», Zentrum für Interdisziplinäre Forschung, Bielefeld, Germania, 14-17 maggio.

Newell, A. (1973), *Production Systems: Models of Control Structures*, in *Visual Information Processing*, a cura di W. G. Chase, Academic Press, New York, pp. 463-526.

— (1982), *The Knowledge Level*, «Artificial Intelligence», 18, pp. 81-132.

— (1988), *The Intentional Stance and the Knowledge Level*, «Behavioral and Brain Sciences», 11, pp. 520-522.

— (1990), *Unified Theories of Cognition*, Harvard University Press, Cambridge, MA.

Newell, A., Rosenbloom, P.S. e Laird, J.E. (1989), *Symbolic Architectures for Cognition*, in *Foundations of Cognitive Science*, a cura di M. Posner, MIT Press, Cambridge, MA. pp. 93-132.

Nielsen, T.I. (1963), *Volition: A New Experimental Approach*, «Scandinavian Journal of Psychology», 4, pp. 225-230.

Nilsson, N. (1984), *Shakey the Computer*, SRI Tech Report, SRI International, Menlo Park.

Norman, D.A. e Shallice, T. (1980), *Attention to Action: Willed and Automatic Control of Behavior*, Center for Human Information Processing (Technical Report n. 99), ristampato con revisioni in *Consciousness and Self-Regulation*, a cura di R. J. Davidson, G. E. Schwartz e D. Shapiro, Plenum Press, New York 1986.

— (1985), *Attention to Action*, in *Consciousness and Self-Regulation*, a cura di T. Shallice, Plenum Press, New York.

Nottebohm, F. (1984), *Birdsong as a Model in Which to Study Brain Processes Related to Learning*, «Condor», 86, pp. 227-236.

Oakley, D. A. (a cura di) (1985), *Brain and Mind*, Methuen, Londra e New York.

Ornstein, R. e Thompson, R. F. (1984), *The Amazing Brain*, Houghton Mifflin, Boston (trad. it. *Il cervello e le sue meraviglie*, Rizzoli, Milano 1987).

Pagels, H. (1988), *The Dreams of Reason: The Computer and the Rise of the Sciences of Complexity*, Simon & Schuster, New York.

Papert, S. (1988), *One AI or Many?*, «Daedalus», inverno, pp. 1-14.

Parfit, D. (1984), *Reasons and Persons*, Clarendon Press, Oxford (trad. it. *Ragioni e persone*, Il Saggiatore, Milano 1989).

Pears, D. (1984), *Motivated Irrationality*, Clarendon Press, Oxford.

Penfield, W. (1958), *The Excitable Cortex in Conscious Man*, Liverpool University Press, Liverpool.

Penrose, R. (1989), *The Emperor's New Mind*, Oxford University Press, Oxford (trad. it. di L. Sosio, *La mente nuova dell'imperatore*, Rizzoli, Milano 1992).

Perlis (1991), *Intentionality and Defaults*, in *Reasonings Agents in a Dyna-*

mic World, a cura di K. M. Ford e P. J. Hayes, JAI Press, Greenwich.
Perry, J. (1979), *The Problem of the Essential Indexical*, «Nous», 13, pp. 3-21.
Pinker, S. e Bloom, P. (1990), *Natural Language and Natural Selection*, «Behavioral and Brain Sciences», 13, pp. 707-784.
Pollatsek, A., Rayner, K. e Collins, W. E. (1984), *Integrating Pictorial Information Across Eye Movements*, «Journal of Experimental Psychology: General», 113, pp. 426-442.
Pöppel, E. (1985), *Grenzen des Bewusstseins*, Deutsche Verlags-Anstal, Stuttgart.
— (1988) (traduzione di Pöppel, 1985), *Mindworks: Time and Conscious Experience*, Harcourt Brace Jovanovich, New York.
Popper, K. R. e Eccles, J. C. (1977), *The Self and Its Brain*, Springer Verlag, Berlino (trad. it. di B. Continenza e G. Mininni, *L'Io e il Suo Cervello*, A. Armando, Roma 1981).
Powers, L. (1978), *Knowledge by Deduction*, «Philosophical Review», 87, pp. 337-371.
Putnam, H. (1965), *Brains and Behavior*, in *Analytical Philosophy*, seconda serie, a cura di R. J. Butler, Blackwell, Oxford, pp. 1-19.
— (1988), *Much Ado About Not Very Much*, «Daedalus», 117, inverno, ristampato in Graubard (1988).
Pylyshyn, Z. (1979), *Do Mental Events Have Durations?*, «Behavioral and Brain Sciences», 2, pp. 227-278.
Quine, W. V. O. (1969), *Natural Kinds*, in *Ontological Relativity and Other Essays*, Columbia University Press, New York, pp. 114-138 (trad. it. di M. Leonelli, *La relatività ontologica ed altri saggi*, Armando, Roma 1986).
Ramachandran, V. S. (1985), editoriale in «Perception», 14, pp. 97-103.
— (1991) *2-D or not 2-D: That Is the Question*, in *The Artful Brain*, a cura di R. L. Gregory, J. Harris, P. Heard, D. Rose e C. Cronly-Dillon, Oxford University Press, Oxford.
Ramachandran, V. S. e Gregory, R. L., inviato a «Nature», *Perceptual Filling in of Artificially Induced Scotomas in Human Vision.*
Ramsey, W., Stich, S. e Rumelhart, D. (a cura di) (1991), *Philosophy and Connectionist Theory*, Erlbaum, Hillsdale.
Raphael, B. (1976), *The Thinking Computer: Mind Inside Matter*, Freeman, San Francisco.
Reddy, D. R., Erman, L. D., Fennel, R. D. e Neely, R. B. (1973), *The HEARSAY-II Speech Understanding System: An Example of the Recognition Process*, «Proceedings of the International Joint Conference on Artificial Intelligence», Stanford, pp. 185-194.
Reingold, E. M. e Merikle, P. M. (1990), *On the Interrelatedness of Theory and Measurement in the Study of Unconscious Processes*, «Mind and Language», 5, pp. 9-28.
Reisberg, D. e Chambers, D. (di prossima pubblicazione), *Neither*

Pictures nor Propositions: What Can We Learn from a Mental Image?, «Canadian Journal of Psychology».
Richards, R. J. (1987), *Darwin and the Emergence of Evolutionary Theories of Mind and Behavior*, Chicago University Press, Chicago.
Ristau, C. (1991), *Cognitive Ethology: The Minds of Other Animals: Essays in Honor of Donald R. Griffin*, Erlbaum, Hillsdale.
Rizzolati, G., Gentilucci, M. e Matelli, M. (1985), *Selective Spatial Attention: One Center, One Circuit, or Many Circuits?*, in *Attention and Performance XI*, a cura di M. I. Posner e O. S. M. Marin, Erlbaum, Hillsdale.
Rorty, R. (1970), *Incorrigibility as the Mark of the Mental*, «Journal of Philosophy», 67, pp. 399-424.
— (1982a), *Contemporary Philosophy of Mind*, «Synthese», 53, pp. 323-348.
— (1982b), *Comments on Dennett*, «Synthese», 53, pp. 181-187.
Rosenbloom, P. S., Laird, J. E. e Newell, A. (1987), *Knowledge-Level Learning in Soar*, «Proceedings of AAAI», Morgan Kaufman, Los Altos.
Rosenthal, D. (1986), *Two Concepts of Consciousness*, «Philosophical Studies», 49, pp. 329-359.
— (1989), *Thinking That One Thinks*, ZIF Report n. 11, Research Group on Mind and Brain, Perspectives in Theoretical Psychology and the Philosophy of Mind, Zentrum für Interdisziplinäre Forschung, Bielefeld, Germania.
— (1990a), *Why Are Verbally Expressed Thoughts Conscious?*, ZIF Report n. 32, Zentrum für Interdisziplinäre Forschung, Bielefeld, Germania.
— (1990b), *A Theory of Consciousness*, ZIF Report n. 40, Zentrum für Interdisziplinäre Forschung, Bielefeld, Germania. .
Rozin, P. (1976), *The Evolution of Intelligence and Access to the Cognitive Unconscious*, «Progress in Psychobiology and Physiological Psychology», 6, pp. 245-280.
— (1982), *Human Food Selection: The Interation of Biology, Culture and Individual Experience*, in *The Psychobiology of Human Food Selection*, a cura di L. M. Barker, Avi Publishing Company, Westport.
Rozin, P. e Fallon, A. E. (1987), *A Perspective on Disgust*, «Psychological Review», 94, pp. 23-47.
Russell, B. (1927), *The Analysis of Matter*, Allen and Unwin, Londra (trad. it. di L. Pavolini, *L'analisi della materia*, Longanesi, Milano 1964).
Ryle, G. (1949), *The Concept of Mind*, Hutchinson, Londra (trad. it. di F. Rossi-Landi, *Lo spirito come comportamento*, Einaudi, Torino 1955).
— (1979), *On Thinking*, a cura di K. Kolenda, Rowman and Littlefield, Totowa (trad. it. di G. Melilli Ramoino, *Pensare pensieri*, Armando, Roma 1990).
Sacks, O. (1985), *The Man Who Mistook His Wife for His Hat*, Summit

Books, New York (trad. it. di C. Morena, *L'uomo che scambiò sua moglie per un cappello*, Adelphi, Milano 1986).
— (1989), *Seeing Voices*, University of California Press, Berkeley (trad. it. di C. Sborgi, *Vedere Voci*, Adelphi, Milano 1990).
Sandeval, E. (1991), *Towards a Logic of Dynamic Frames*, in *Reasonings Agents in a Dynamic World*, a cura di K. M. Ford e P. J. Hayes, JAI Press, Greenwich.
Sanford, D. (1975), *Infinity and Vagueness*, «Philosophical Review», 84, pp. 520-535.
Sartre, J.-P. (1943), *L'Etre et le Néant*, Gallimard, Parigi (trad. it. *L'essere e il nulla*, Il Saggiatore, Milano 1984).
Schank, R. (1991), *Tell Me a Story*, Scribners, New York.
Schank, R. e Abelson, R. (1977), *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*, Erlbaum, Hillsdale.
Schull, J. (1990), *Are Species Intelligent?*, «Behavioral and Brain Sciences», 13, pp. 63-108.
Searle, J. (1980), *Minds, Brains, and Programs*, «Behavioral and Brain Sciences», 3, pp. 417-458 (trad. it. in *Menti, cervelli e programmi: un dibattito sull'intelligenza artificiale*, a cura di G. Tonfoni, CLUP-CLUED, Milano 1984; oppure in Hofstadter e Dennett (1981)).
— (1982), *The Myth of the Computer: An Exchange*, «New York Review of Books», 24 giugno, pp. 56-57.
— (1983), *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge (trad. it. *Della intenzionalità: un saggio di filosofia della conoscenza*, Bompiani, Milano 1985).
— (1984), *Panel Discussion: Has Artificial Intelligence Research Illuminated Human Thinking?*, in *Computer Culture: The Scientific, Intellectual, and Social Impact of the Computer*, «Annals of the New York Academy of Sciences», 426.
— (1988a), *Turing the Chinese Room*, in *Synthesis of Science and Religion, Critical Essays and Dialogues*, a cura di T. Singh, Bhaktivedenta Institute, San Francisco 1988.
— (1988b), *The Realistic Stance*, «Behavioral and Brain Sciences», 11, pp. 527-529.
— (1990a), *Consciousness, Explanatory Inversion, and Cognitive Science*, «Behavioral and Brain Sciences», 13, pp. 585-642.
— (1990b), *Is the Brain's Mind a Computer Program*, «Scientific American», 262, pp. 26-31.
Selfridge, O. (1959), *Pandemonium: A Paradigm for Learning*, Symposium on the Mechanization of Thought Processes, HM Stationery Office, Londra.
— (non pubblicato), *Tracking and Trailing*.
Sellars, W. (1963), *Empiricism and the Philosophy of Mind*, in *Science, Perception and Reality*, Routledge & Kegan Paul, Londra.
— (1981), *Foundations for a Metaphysics of Pure Process*, (the Carus Lectures), «Monist», 64, pp. 3-90.

Shallice, T. (1972), *Dual Functions of Consciousness*, «Psychological Review», 79, pp. 383-393.
— (1978), *The Dominant Action System: An Information-Processing Approach to Consciousness*, in *The Stream of Consciousness*, a cura di K. S. Pope e J. L. Singer, Plenum, New York, pp. 148-164.
— (1988), *From Neuropsychology to Mental Structure*, Cambridge University Press, Cambridge.
Sharpe, T. (1977), *The Great Pursuit*, Secker and Warburg, Londra (trad. it. di C. Brera, *La grande caccia*, Longanesi, Milano 1989).
Shepard, R. N. (1964), *Circularity in Judgments of Relative Pitch*, «Journal of the Acoustical Society of America», 36, pp. 2346-2353.
Shepard, R. N. e Cooper, L. A. (1982), *Mental Images and Their Transformations*, MIT Press/A Bradford Book, Cambridge, MA.
Shepard, R. N. e Metzler, J. (1971), *Mental Rotation of Three-Dimensional Objects*, «Science», 171, pp. 701-703.
Shoemaker, S. (1969), *Time Without Change*, «Journal of Philosophy», 66, pp. 363-381.
— (1975), *Functionalism and Qualia*, «Synthese», 27, pp. 291-315.
— (1981), *Absent Qualia are Impossible – A Reply to Block*, «Philosophical Review», 90, pp. 581-599.
— (1988), *Qualia and Consciousness*, Tufts University Philosophy Department Colloquium.
Siegel, R. K. e West, L. J. (a cura di) (1975), *Hallucinations: Behavior, Experience and Theory*, Wiley, New York.
Simon, H. A. e Kaplan, C. A. (1989), *Foundations of Cognitive Science*, in *Foundations of Cognitive Science*, a cura di Posner, MIT Press, Cambridge, MA.
Smolensky, P. (1988), *On the Proper Treatment of Connectionism*, «Behavioral and Brain Sciences», 11, 1-74.
Smullyan, R. M. (1981), *An Epistemological Nightmare*, in Hofstadter e Dennett (1981), pp. 415-427 (trad. it. *Un incubo epistemologico*, pp. 403-413).
Smythies, J. R. (1954), *Analysis of Projection*, «British Journal of Philosophy of Science», 5, pp. 120-133.
Snyder, D. M. (1988), *On the Time of a Conscious Peripheral Sensation*, «Journal of Theoretical Biology», 130, pp. 253-254.
Sperber, D. e Wilson, D. (1986), *Relevance: A Theory of Communication*, Harvard University Press, Cambridge, MA.
Sperling, G. (1960), *The Information Available in Brief Visual Presentations*, «Psychological Monographs», 74, n. 11.
Sperry, R. W. (1977), *Forebrain Commissurotomy and Conscious Awareness*, «The Journal of Medicine and Philosophy», 2, pp. 101-126.
Spillman, L. e Werner, J. S. (1990), *Visual Perception: The Neurophysiological Foundations*, Academic Press, San Diego.
Spinoza, B. (1661), *Tractatus de intellectus emendatione* (trad. it. di E. De Angelis, *Emendazione dell'intelletto*, Boringhieri, Torino 1962).
Stafford, S. P. (1983), *On The Origin of the Intentional Stance*, Tufts

University Working Paper in Cognitive Science, CCM 83-1.
Stalnaker, R. (1984), *Inquiry*, MIT Press/A Bradford Book, Cambridge, MA.
Stix, G. (1991), *Reach Out*, «Scientific American», 264, p. 134.
Stoerig, P. e Cowey, A. (1990), *Wavelength Sensitivity in Blindsight*, «Nature», 342, pp. 916-918.
Stoll, C. (1989), *The Cuckoo's Egg: Tracking a Spy Through the Maze of Computer Espionage*, Doubleday, New York (trad. it. di E. De Angelis, *L'uovo del cuculo*, Sperling & Kupfer, Milano 1990).
Straight, H. S. (1976), *Comprehension versus Production in Linguistic Theory*, «Foundations of Language», 14, pp. 525-540.
Stratton, G. M. (1896), *Some Preliminary Experiments on Vision Without Inversion of the Retinal Image*, «Psychology Review», 3, pp. 611-617.
Strawson, G. (1989), *Red and «Red»*, «Synthese», 78, pp. 193-232.
Strawson, P. F. (1962), *Freedom and Resentment*, «Proceedings of the British Academy», ristampato in *Studies in the Philosophy of Thought and Action*, a cura di P. F. Strawson, Oxford University Press, Oxford 1968.
Taylor, D. M. (1966), *The Incommunicability of Content*, «Mind», 75, pp. 527-541.
Thompson, D'Arcy W. (1917), *On Growth and Form*, Cambridge University Press, Cambridge (trad. it. di J. T. Bonner, *Crescita e forma*, Boringhieri, Torino 1969).
Thompson, E., Palacios, A. e Varela, F. (in stampa), *Ways of Coloring*, «Behavioral and Brain Sciences».
Tranel, D. e Damasio, A. R. (1988), *Non-conscious Face Recognition in Patients with Face Agnosia*, «Behavioral Brain Research», 30, pp. 235-249.
Tranel, D., Damasio, A. R. e Damasio, H. (1988), *Intact Recognition of Facial Expression, Gender, and Age in Patients with Impaired Recognition of Face Identity*, «Neurology», 38, pp. 690-696.
Treisman, A. (1988), *Features and Objects: The Fourteenth Bartlett Memorial Lecture*, «Quarterly Journal of Experimental Psychology», 40A, pp. 201-237.
Treisman, A. e Gelade, G. (1980), *A Feature-integration Theory of Attention*, «Cognitive Psychology», 12, pp. 97-136.
Treisman, A. e Sato, S. (1990), *Conjunction Search Revisited*, «Journal of Experimental Psychology: Human Perception and Performance», 16, pp. 459-478.
Treisman, A. e Souther, J. (1985), *Search Asimmetry: A Diagnostic for Preattentive Processing of Separable Features*, «Journal of Experimental Psychology: General», 114, pp. 285-310.
Turing, A. (1950), *Computing Machinery and Intelligence*, «Mind», 59, pp. 433-460 (trad. it. *Calcolatori e Intelligenza*, in Hofstadter e Dennett [1981], pp. 61-75).
Tye, M. (1986), *The Subjective Qualities of Experience*, «Mind», 95, pp. 1-17.

Uttal, W. R. (1979), *Do Central Nonlinearities Exist?*, «Behavioral and Brain Sciences», 2, p. 286.
Van der Waals, H. G. e Roelofs, C. O. (1930), *Optische Scheinbewegung*, «Zeitschrift für Psychologie und Physiologie des Sinnesorgane», 114, pp. 241-288, 115 (1931), pp. 91-190.
Van Essen, D. C. (1979), *Visual Areas of the Mammalian cerebral Cortex*, «Annual Review of Neuroscience», 2, pp. 227-263.
van Gulick, R. (1988), *Consciousness, Intrinsic Intentionality, and Self-understanding Machines*, in Marcel e Bisiach (1988), pp. 78-100.
— (1989), *What Difference Does Consciousness Make?*, «Philosophical Topics», 17, pp. 211-230.
— (1990), *Understanding the Phenomenal Mind: Are We All Just Armadillos?*, presentato alla conferenza «The Phenomenal Mind – How Is It Possible and Why Is It Necessary?», Zentrum für Interdisziplinäre Forschung, Bielefeld, Germania, 14-17 maggio.
van Tuijl, H. F. J. M. (1975), *A New Visual Illusion: Neonlike Color Spreading and Complementary Color Induction between Subjective Contours*, «Acta Psychologica», 39, pp. 441-445.
Vendler, Z. (1972), *Res Cogitans*, Cornell University Press, Ithaca.
— (1984), *The Matter of Minds*, Clarendon Press, Oxford.
von der Malsburg, C. (1985), *Nervous Structures with Dynamical Links*, «Berichte der Bunsen-Gesellschaft für Physikalische Chemie», 89, pp. 703-710.
von Uexküll, J. (1909), *Umwelt und Innenwelt der Tiere*, Jena, Berlino.
Vosberg, R., Fraser, N. e Guehl, J. (1960), *Imagery Sequence in Sensory Deprivation*, «Archives of General Psychiatry», 2, pp. 356-357.
Walton, K. (1973), *Pictures and Make Believe*, «Philosophical Review», 82, pp. 283-319.
— (1978), *Fearing Fiction*, «Journal of Philosophy», 75, pp. 6-27.
Warren, R. M. (1970), *Perceptual Restoration of Missing Speech Sounds*, «Science», 167, pp. 392-393.
Wasserman, G. S. (1985), *Neural/Mental Chronometry and Chronotheology*, «Behavioral and Brain Sciences», 8, pp. 556-557.
Weiskrantz, L. (1986), *Blindsight: A Case Study and Implications*, Oxford University Press, Oxford.
— (1988), *Some Contributions of Neuropsychology of Vision and Memory to the Problem of Consciousness*, in Marcel e Bisiach (1988), pp. 183-199.
— (1989), «Panel discussion on consciousness», European Brain and Behavior Society, Torino, settembre 1989.
— (1990), *Outlooks for Blindsight: Explicit Methodologies for Implicit Processes* (the Ferrier Lecture), «Proceedings of the Royal Society London», B 239, pp. 247-278.
Welch, R. B. (1978), *Perceptual Modification: Adapting to Altered Sensory Environments*, Academic Press, New York.
Wertheimer, M. (1912), *Experimentelle Studien über das Sehen von Bewegung*, «Zeitschrift für Psychologie», 61, pp. 161-265.

White, S. L. (1986), *The Curse of the Qualia*, «Synthese», 68, pp. 333-68.
Whiten, A. e Byrne, R. (1988), *Toward the Next Generation in Data Quality: A New Survey of Primate Tactical Deception*, «Behavioral and Brain Sciences», 11, pp. 267-273.
Wiener, N. (1948), *Cybernetics: or Control and Communication in the Animal and the Machine*, Technology Press, Cambridge (trad. it. di O. Beghelli, *La cibernetica*, Bompiani, Milano 1953).
Wilkes, K. V. (1988), *Real People*, Oxford University Press, Oxford.
Wilsson, L. (1974), *Observations and Experiments on the Ethology of the European Beaver*, «Viltrevy, Swedish Wildlife», 8, pp. 115-266.
Winograd, T. (1972), *Understanding Natural Language*, Academic Press, New York.
Wittgenstein, L. (1953), *Philosophische Untersuchungen*, Blackwell, Oxford (trad. it. di M. Trinchero, *Ricerche filosofiche*, Einaudi, Torino 1967).
Wolfe, J. M. (1990), *Three Aspects of the Parallel Guidance of Visual Attention*, «Proceedings of the Cognitive Science Society», Erlbaum, Hillsdale, pp. 1048-1049.
Yonas, A. (1981), *Infants' Responses to Optical Information for Collision*, in *Development of Perception: Psychobiological Perspectives*, Vol. 2: *The Visual System*, a cura di R. N. Aslin, J. R. Alberts e M. R. Peterson, Academic Press, New York.
Young, J. Z. (1965a), *The Organization of a Memory System*, «Proceedings Royal Society London [Biology]», 113, pp. 285-320.
— (1965b), *A Model of the Brain*, Clarendon, Oxford (trad. it. di F. Ciafaloni, *Un modello del cervello*, Einaudi, Torino 1974).
— (1979), *Learning as a Process of Selection*, «Journal of the Royal Society of Medicine», 72, pp. 801-804.
Zajonc, R. e Markus, H. (1984), *Affect and Cognition: The Hard Interface*, in *Emotion, Cognition and Behavior*, a cura di C. Izard, J. Kagan e R. Zajonc, Cambridge University Press, Cambridge, pp. 73-102.
Zeki, S. M. e Shipp, S. (1988), *The Functional Logic of Cortical Connections*, «Nature», 335, pp. 311-317.
Zihl, J. (1980), *«Blindsight»: Improvement of Visually Guided Eye Movements by Systematic Practice in Patients with Cerebral Blindness*, «Neuropsychologica», 18, pp. 71-77.
— (1981), *Recovery of Visual Functions in Patients with Cerebral Blindness*, «Experimental Brain Research», 44, pp. 159-169.

Avvertenza. Il saggio di Dennett e Kinsbourne, *Time and the Observer*, è stato pubblicato nel 1992 [15 (2), pp. 183-234]; quello di Marcel, *Slippage in the Unity of Consciousness*, nel 1993 in Bock, G.R., Marsh, J. (eds.), *Experimental and Theoretical Studies of Consciousness*, Ciba Foundation Symposium 174, John Wiley & Sons, Chichester, pp. 168-180; quello di Ramachandran e Gregory, *Perceptual Filling in of Artificially Induced Scotomas in Human Vision*, nel 1991 [350 (6320), pp. 699-702]; quello di Reisberg e Chambers, *Neither Pictures nor Propositions*, nel 1991 [45, pp. 288-302]; quello di Thompson, Palacios e Varela, *Ways of Coloring*, nel 1992 [15, pp. 1-74]. Lo studio di Humphrey, *A History of the Mind*, è stato pubblicato nel 1992 presso Simon & Schuster, New York, e tradotto in italiano nel 1998, presso Instar Libri, Torino.

POSTFAZIONE
LA TEORIA DELLA COSCIENZA DI DANIEL C. DENNETT: GUIDA AGLI APPROFONDIMENTI BIBLIOGRAFICI

*di Massimo Marraffa**

Alla fine del XX secolo, dopo un lungo periodo di eclissi, i fenomeni rubricati sotto il termine «coscienza» hanno riacquistato la dignità di oggetti d'indagine scientifica. Nel quadro di questa riabilitazione sono state proposte varie teorie *cognitiviste* della coscienza, teorie, cioè, che si propongono di spiegare i fenomeni della coscienza avvalendosi dell'apparato computazionale e rappresentazionale della scienza cognitiva[1]. A questa opera di costruzione teorica anche alcuni filosofi «orientati empiricamente» hanno dato il loro contributo; un apporto che, almeno nel caso del volume che il lettore ha appena finito di leggere (d'ora in poi *Coscienza*), è stato estremamente importante.

In questa postfazione cercheremo di far emergere alcune ragioni di questa importanza attraverso una breve ricognizione dell'intricato dibàttito scaturito dalla prima apparizione di *Coscienza*, diciotto anni or sono[2].

La base di tutta la riflessione filosofica di Dennett è una strategia per lo studio della mente esposta per la prima volta in *Content and Consciousness*[3] (la tesi di dottorato scritta sotto la guida di Gilbert Ryle[4]). Tale strategia consiste in due passi: *prima* si costruisce una teoria del contenuto che sia *indipendente* dalla coscienza e più *fondamentale* di essa[5]; e quindi, su queste basi, si procede a sviluppare una teoria della coscienza, concepita come «un fenomeno avanzato o derivato» e non già, idealisticamente, come «il fondamento di tutta l'intenzionalità, di tutta la mentalità»[6]. Dopo la prima formulazione in *Content and Consciousness*, la strategia «prima il contenuto, poi la coscienza» è stata oggetto di almeno due importanti messe a punto: la prima in *Brainstorms*; la seconda in *The In-*

* *Un ringraziamento a Erica Cosentino e Michele Di Francesco per alcuni utili commenti su una versione precedente di questo scritto.*

tentional Stance e in *Coscienza*. In *Kinds of Minds* Dennett ha quindi cercato di offrire una sintesi unitaria delle sue teorie del contenuto e della coscienza, mentre in *Darwin's Dangerous Idea* ha provato a delinearne i fondamenti evoluzionistici[7].

Respinta la visione idealistica della coscienza come entità dotata di un carattere primario e fondante rispetto al resto della vita psichica, Dennett scorge nelle scienze della mente gli strumenti che consentono di comprendere la fenomenologia della prima persona come l'autopresentarsi di eventi che hanno luogo nell'inconscio computazionale. Nel suo idioma, l'*auto*fenomenologia deve far posto all'*etero*fenomenologia. Centro focale del cap. 4 di *Coscienza*[8], il metodo dell'eterofenomenologia è stato variamente descritto come una banale ridescrizione di pratiche familiari, una riformulazione scarsamente originale di Husserl (o il suo tradimento), una proposta rivoluzionaria circa il modo in cui studiare la coscienza, una forma di cripto-comportamentismo, un esagerato attacco al senso comune – e l'elenco potrebbe continuare[9]. Di fronte al proliferare incontrollato delle interpretazioni, Dennett si è sforzato a più riprese di chiarire natura e scopi del metodo eterofenomenologico[10].

Il rifiuto dell'autofenomenologia è l'esito dello scetticismo che Dennett nutre nei riguardi dell'introspezione[11]. Il *locus classicus* dell'anti-introspezionismo contemporaneo è *Telling more than we can know: Verbal reports on mental processes* di Richard E. Nisbett e Timothy D. Wilson[12]. In questa celebre rassegna della letteratura sull'attribuzione causale e la dissonanza cognitiva sono descritti numerosi esperimenti in cui i soggetti invece di accedere introspettivamente alle *motivazioni* reali (ossia alle cause) della loro condotta nell'esperimento, «confabulano», confezionando storie causali che hanno poco o nulla a che fare con le storie reali[13]. Questi dati della psicologia sociale sono congruenti con quelli ottenuti da alcuni neuropsicologi e psicologi dello sviluppo[14]. Dennett generalizza questi risultati empirici sostenendo che *tutti* i resoconti introspettivi sono produzioni verbali di utili finzioni[15]. Ciò malgrado, vi è chi ha fatto notare che l'anti-introspezionismo di Dennett non è sufficientemente radicale nella misura in cui concede al soggetto ancora una qualche forma di autorità epistemica nell'esplicitazione verbale dell'esperienza cosciente[16].

Una difesa radicale dell'autofenomenologia è venuta da quei filosofi che ritengono che l'unico senso legittimo di coscienza sia quello *fenomenico*[17]. In quest'ottica, la coscienza è inderivata e inseparabile dall'intenzionalità, nonché essenziale per l'esistenza di

quest'ultima[18]. Per contrastare questa posizione Dennett ha originalmente integrato due mosse teoriche disponibili nella letteratura sulla coscienza fenomenica.

La prima mossa consiste nel ridefinire la coscienza fenomenica in termini funzionali e rappresentazionali, ossia come «coscienza d'accesso»[19]. L'esempio più importante di questo approccio è offerto da William Lycan. Per questo studioso la coscienza è un sistema di automonitoraggio del cervello, per cui gli stati di coscienza sono esperienze di (altri) stati mentali[20]. (Questa teoria della coscienza è nota come «teoria delle *esperienze* di ordine superiore» e si distingue dalla «teoria dei *pensieri* di ordine superiore» in quanto non fa dipendere la consapevolezza di ordine superiore dall'intervento di un atto di giudizio, ossia da una rappresentazione concettuale dell'esperienza in questione[21].) Nel contesto di questa visione della coscienza come «senso interno», Lycan propone la tesi secondo cui i *qualia* sono contenuti intenzionali, proprietà rappresentate di oggetti rappresentati[22]; e colloca questa teoria rappresentazionalista dei *qualia* in una prospettiva funzionalista, il che, per l'appunto, gli consente di riassorbire pienamente la coscienza fenomenica nella coscienza d'accesso.

La seconda mossa teorica consiste nel negare l'esistenza della coscienza fenomenica o, quanto meno, di quelle fra le sue proprietà che si oppongono alla naturalizzazione. Gli studiosi più radicali respingono totalmente la nozione di coscienza in quanto ritengono che la distinzione cosciente/non-cosciente non consenta di ritagliare la realtà psichica lungo le sue articolazioni reali[23]. Tali teorie asseriscono che la nozione di coscienza è talmente confusa da meritare l'eliminazione e la sostituzione da parte di concetti e distinzioni che rispecchiano maggiormente la vera natura della mente[24]. Nella maggior parte dei casi, però, l'eliminazionismo non investe l'intera nozione di coscienza ma solo alcune delle sue proprietà. Pertanto, accade spesso che un'eliminazione selettiva sia proposta nel quadro di una teoria positiva di quegli aspetti della coscienza che sono giudicati reali.

Ora, come si è detto, Dennett compie entrambe le mosse sopracitate. Il suo *eliminazionismo* investe solo alcune proprietà della coscienza, come i *qualia*, l'io e il «teatro cartesiano». Nello stesso tempo, il modello delle molteplici versioni è *rappresentazionalista* in quanto analizza la coscienza esclusivamente in termini di relazioni di contenuto; ed è anche, in una certa misura, una *teoria di ordine superiore*, dal momento che considera «l'*io* un aspetto virtuale o emergente della narrazione approssimativamente seriale e coe-

rente che è costruita attraverso il gioco interattivo dei contenuti all'interno del sistema»[25].

Per arrivare al modello delle molteplici versioni Dennett ha impiegato ventidue anni. La prima versione della sua teoria della coscienza si trova in *Content and Consciousness*. Questa è stata poi sviluppata nella seconda parte di *Brainstorms* e in una serie di saggi meno noti. Il contenuto di questi ultimi è stato quindi rielaborato e combinato con nuovi materiali – non ultimi quelli prodotti dalla collaborazione con il neuropsicologo Marcel Kinsbourne[26] –, dando luogo a *Coscienza*. Negli ultimi diciotto anni Dennett ha continuato a sviluppare la sua teoria attraverso una lunga serie di risposte ai critici[27], alcune delle quali sono state raccolte nel recente *Sweet Dreams*[28].

Alcuni critici descrivono Dennett come «un bersaglio in movimento»[29]; altri ritengono che non si sia mosso di un millimetro da *Content and Consciousness*. Un'interpretazione «continuista» è quella proposta da Matthew Elton, secondo cui la teoria della coscienza di Dennett sarebbe rimasta fortemente stabile tra il 1969 e il 1991. Nel cap. 6 di *Content and Consciousness* Dennett traccia una distinzione tra due modi in cui un sistema intenzionale può essere consapevole di stati di cose:

> la consapevolezza$_1$, l'eccentrico tipo di coscienza che noi esseri umani godiamo in virtù della capacità di formulare resoconti introspettivi (verbali); e la consapevolezza$_2$, la mera capacità di rispondere in modo adeguato agli stimoli, una capacità di cui godono in egual misura le api e i termostati[30].

Ora, sostiene Elton, il nocciolo duro della teoria della coscienza di Dennett è l'idea che la consapevolezza$_1$ di uno stato di cose è condizione necessaria e sufficiente dell'avere coscienza di tale stato[31]. Questo è un risultato dell'analisi concettuale al livello *personale*; un'acquisizione stabile che è poi diversamente declinata in *Brainstorms* e in *Coscienza* col mutare delle ipotesi circa i meccanismi *subpersonali* alla base della «capacità di formulare resoconti introspettivi (verbali)»[32].

Uno degli sviluppi subpersonali più importanti della teoria della coscienza di Dennett è senza dubbio la sua soluzione del problema della «temporalizzazione» (*timing*) e dell'ordinamento degli oggetti nel flusso di coscienza. Questa soluzione comincia a prendere forma nel saggio *Are Dreams Experiences?*[33], è approfondita in *Two Approaches to Mental Imagery*[34] e in *Quining qualia*[35], e raggiunge quindi la sua piena articolazione nei capp. 5 e 6 di *Coscienza*[36]. Qui,

si rammenterà, le anomalie temporali della coscienza (il phi colorato, il coniglio cutaneo, il riferimento all'indietro nel tempo e il ritardo soggettivo della coscienza dell'intenzione) sono la base empirica del modello delle molteplici versioni. In tutti questi casi, sostiene Dennett, il materialismo cartesiano non ci permette di discriminare tra la storia orwelliana e quella staliniana; tutto quello che abbiamo è «una differenza che non fa differenza»[37]. Ciò conferisce plausibilità al modello delle molteplici versioni, dal momento che nega l'esistenza di una linea il cui attraversamento sancisce la fine dell'elaborazione inconscia o preconscia e l'inizio della valutazione cosciente[38].

Sulla natura e la plausibilità di questa tesi si è acceso un intenso dibattito[39]. In particolare, le critiche si sono appuntate sulla posizione verificazionista di Dennett per cui, data l'impossibilità di discriminare fra la revisione orwelliana e quella staliniana, non vi sarebbe al riguardo alcuna «verità» (*fact of the matter*)[40]. Un'altra questione molto controversa è se effettivamente non vi sia alcuna differenza, nemmeno in linea di principio, tra le due revisioni[41]. Per esempio, Ned Block ha sostenuto che questa tesi poggia interamente su una posizione antirealista nei riguardi della coscienza fenomenica, che però, a suo dire, Dennett non argomenta[42]. A questa obiezione Dennett e Kinsbourne hanno risposto respingendo l'etichetta di eliminazionisti e indicando nella scala temporale microscopica la vera ragione per cui è impossibile discriminare tra la revisione orwelliana e quella staliniana[43].

Ma allora, vien da chiedersi, Dennett è o non è un eliminazionista nei confronti della coscienza fenomenica? Secondo Susan Schneider, la risposta dipende dal significato che si attribuisce al termine «quale»[44]. In *Quining qualia* e in *Coscienza*, Dennett definisce i *qualia* come proprietà che sono (i) ineffabili, (ii) intrinseche, (iii) private e (iv) immediatamente accessibili; e completando la demolizione del privato cartesiano avviata da Wittgenstein, Ryle, Austin e Sellars, il filosofo afferma che non esiste alcunché in grado di soddisfare la descrizione (i)-(iv). A sostegno di questa tesi *eliminazionista*, Dennett ha cercato di confutare tutti i principali argomenti in favore dei *qualia* – in particolare, gli argomenti della conoscenza, dei *qualia* inversi e dei *qualia* assenti[45]. Tuttavia, ha sostenuto Schneider, la tesi eliminazionista non sembrerebbe incompatibile con la realtà dei *qualia* presi nel senso più generale (e più comune) di proprietà che costituiscono «ciò che si prova» (*what it's like*) quando si ha un'esperienza cosciente[46]. Così intesi, i *qualia* potrebbero essere un osso meno duro per l'approccio rappresentazionalista.

In ossequio a una lunga tradizione in filosofia e nelle scienze sociali, Dennett concepisce il linguaggio come il medium di *tutto* il pensiero concettuale umano. Per il filosofo, la coscienza umana non è un sistema biologico ma una macchina virtuale «neumanniana». Questa macchina è il prodotto di comportamenti appresi («buoni trucchi» o «memi»), che hanno riprogrammato i nostri cervelli biologici[47]. Questi comportamenti hanno natura *linguistica*: la coscienza è acquisita attraverso forme di «autostimolazione» linguistica, come ad esempio parlare a se stessi, producendo, ripetendo e riorganizzando enunciati nel soliloquio manifesto o silenzioso (qui Dennett può far leva sulle ricerche sul «linguaggio interno» ispirate da Vygotskij). Questo flusso di verbalizzazione interna ha trasformato l'attività cerebrale dei primi ominidi, facendo sì che la loro architettura parallela simulasse il comportamento di un elaboratore seriale, operante sugli enunciati delle lingue naturali (la «macchina joyciana»). In quest'ottica, la mente concettuale è una costruzione sociale, che si costituisce grazie all'assorbimento dei memi da parte della cultura circostante[48]. Si tratta, dunque, di una versione molto forte dell'idea che il linguaggio svolge un ruolo *costitutivo* nel pensiero; e data la radicalità, la tesi non poteva andare esente da critiche[49].

Ma che tipo di programma gira sulla macchina joyciana? In *Coscienza* (p. 214) questo programma è il *Pandemonio* di Oliver Selfridge, «un modello di architettura computazionale non gerarchica e competitiva», precursore del connessionismo[50]. Il *Pandemonio* possiede due caratteristiche importanti per Dennett: primo, la sua architettura è a parallelismo elevato, e quindi non ha bisogno di postulare un esecutivo centrale; secondo, il *Pandemonio* impiega la scomposizione funzionale ricorsiva, un requisito indispensabile dal momento che nel teatro cartesiano si annida un super-*homunculus* di cui sbarazzarsi[51].

Più di recente, Dennett ha esaminato anche un'altra architettura funzionale compatibile col modello delle molteplici versioni: la teoria della coscienza come «spazio di lavoro globale» proposta da Bernard J. Baars[52]. Secondo questa influente teoria, «la coscienza è realizzata da una società distribuita di specialisti, munita di una memoria di lavoro denominata "spazio di lavoro globale", i cui contenuti possono essere trasmessi al sistema nel suo complesso»[53]. Dennett ha rilevato vari punti in comune tra questa teoria e il suo modello delle molteplici versioni. Primo fra tutti, il ricorso al parallelismo elevato:

il punto chiave [...] sta nel fatto che è l'accessibilità dei demoni specialisti *l'uno all'altro* (e non a qualche immaginario alto Dirigente o Ego centrale) che in linea di principio può spiegare gli spettacolari incrementi di competenza cognitiva che associamo alla coscienza: [...] Quest'idea era al centro anche di quello che ho chiamato il modello delle molteplici versioni[54].

E l'affinità con la teoria dello spazio di lavoro globale si accentua con la nuova metafora che Dennett affianca a quella delle molteplici versioni: la «fama nel cervello» o «celebrità cerebrale»[55]. Al pari della «fama», la coscienza non è una proprietà intrinseca dei processi cerebrali, ma è più simile al «peso politico»: gli eventi coscienti sono quelli che hanno effetti diffusi nel cervello. E, osserva Dennett, la teoria dello spazio di lavoro globale ci consente di ipotizzare che tale «peso» sia realizzato «dalla 'riverberazione' in un 'ciclo di amplificazione prolungata' dei contenuti vincenti»[56].

Tuttavia, si è fatto osservare[57], i sostenitori della teoria della coscienza come spazio di lavoro globale fanno uso massiccio di metafore *teatrali*, in cui gli eventi coscienti si verificano nel «teatro della coscienza» e sullo «schermo della coscienza»[58]. Se il ricorso a simili metafore dovesse rivelarsi qualcosa di più di un artificio espositivo estrinseco al modello, quest'ultimo sarebbe – con buona pace di Dennett – una riproposizione di alcuni aspetti del teatro cartesiano[59].

Roma, ottobre 2008

Note

1. Le guide più recenti e aggiornate al dibattito filosofico e scientifico sulla coscienza sono: *The Cambridge Handbook of Consciousness*, a cura di P.D. Zelazo, M. Moscovitch e E. Thompson, Cambridge UP, Cambridge 2007; *The Blackwell Companion to Consciousness*, a cura di M. Velmans e S. Schneider, Blackwell, Oxford 2007; *The Oxford Companion to Consciousness*, a cura di T. Bayne, A. Cleeremans e P. Wilken, Oxford UP, Oxford, in corso di stampa.
2. La più ampia disamina critica del lavoro di Dennett sulla coscienza è il doppio fascicolo di *Philosophical Topics*, 22(1-2), 1994, pp. 505-568, che contiene 14 saggi e una lunga risposta di Dennett. Ma analisi importanti sono contenute anche in *Dennett and his Critics: Demystifying Mind*, a cura di B. Dahlbom, Blackwell, Oxford 1993; *Dennett's Philosophy: A Comprehensive Assessment*, a cura di D. Ross, A. Brook e D. Thompson, MIT Press, Cambridge (MA) 2000 (in particolare, vedi il saggio di Brook, *Judgments and drafts eight years later*); *Daniel Dennett*, a cura di A. Brook e D. Ross, Cambridge UP, Cambridge 2002; J. Symons, *On Dennett*, Wadsworth, Belmont (CA) 2002; Id., *Dennett: un naturalisme en chantier*, Presses Universitaires France, Paris

2005; M. Elton, *Daniel Dennett. Reconciling Science and Our Self-Conception*, Polity Press, Oxford 2003; T. Zawidzki, *Dennett*, Oneworld Publications, Oxford 2007.

3. D.C. Dennett, *Content and Consciousness*, Routledge, London 1969 (trad. it., *Contenuto e coscienza*, Il Mulino, Bologna 1992).
4. Laureatosi in filosofia ad Harvard nel 1963, Dennett si trasferisce ad Oxford per lavorare sotto la supervisione di Gilbert Ryle – il cui libro *The Concept of Mind* (Hutchinson, London 1949; trad. it., *Il concetto di mente*, Laterza, Roma-Bari 2007) gli «era sembrato uno dei pochi libri di filosofia contemporanea degno di nota» (Dennett, *Dove nascono le idee*, Di Renzo, Roma 2006, p. 15). Nel suo *Self-profile* (in *A Companion to the Philosophy of Mind*, a cura di S. Guttenplan, Blackwell, Oxford 1994, p. 237), Dennett descrive Ryle e Quine (con il quale aveva studiato ad Harvard) come i suoi «mentori filosofici». Questa dichiarazione va presa alla lettera dal momento che Dennett è sicuramente «colui che più di ogni altro ha saputo combinare in modo originale la filosofia della mente di Ryle e Wittgenstein con il naturalismo epistemologico di Quine» (S. Nannini, *L'anima e il corpo*, Laterza, Roma-Bari 2002, p. 166).
5. *Coscienza*, p. 509.
6. D.C. Dennett, *Review of J. Searle, "The Rediscovery of the Mind"*, in «Journal of Philosophy», 60(4), 1993, p. 193.
7. D.C. Dennett, *Brainstorms*, MIT Press, Cambridge (MA) 1978 (trad. it., *Brainstorms*, Adelphi, Milano 1991); Id., *The Intentional Stance*, MIT Press, Cambridge (MA) 1987 (trad. it., *L'atteggiamento intenzionale*, Il Mulino, Bologna 1993); Id., *Kinds of Minds: Towards an Understanding of Consciousness*, Weidenfeld & Nicolson, London 1996 (trad. it., *La mente e le menti*, Rizzoli, Milano 2000); Id., *Darwin's Dangerous Idea: Evolution and the Meanings of Life*, Simon & Schuster, New York 1995 (trad. it., *L'idea pericolosa di Darwin. L'evoluzione e i significati della vita*, Bollati Boringhieri, Torino 2004).
8. Il capitolo si basa su precedenti trattazioni del metodo eterofenomenologico: D.C. Dennett, *Two Approaches to Mental Imagery*, in *Brainstorms* cit., cap. 10; Id., *How to study human consciousness empirically, or nothing comes to mind*, in «Synthese», 59, 1982, pp. 159-180; *Beyond belief*, in *Thought and Object: Essays on Intentionality*, a cura di A. Woodfield, Oxford UP, Oxford 1982 (rist. in *The Intentional Stance* cit., pp. 117-202).
9. Vedi D. Carr, *Phenomenology and fiction in Dennett*, in «International Journal of Philosophical Studies», 6(3), 1998, pp. 331-344; D.L. Thompson, *Phenomenology and heterophenomenology: Husserl and Dennett on reality and science*, in *Dennett's Philosophy* cit., pp. 201-218; S. Gallagher, *Phenomenology and experimental design*, in *Trusting the Subject*, a cura di A. Jack e A. Roepstorff, Imprint Academic, Exeter 2003, pp. 85-99; G. Piccinini, *Data from introspective reports: Upgrading from commonsense to science*, ivi, pp. 141-156; e soprattutto i saggi contenuti in «Phenomenology and Cognitive Sciences», 6(1-2), 2007, pp. 1-270, fascicolo interamente dedicato all'eterofenomenologia, a cura di Alva Noë.
10. D.C. Dennett, *Tiptoeing past the covered wagons*, Emory Cognition Project, Report #28, Emory University, aprile 1994, <http://ase.tufts.edu/cogstud/papers/tiptoe.htm>; Id., *Who's on first? Heterophenomenology explained*, in *Trusting the Subject* cit., pp. 19-30; *A third-person approach to consciousness*, in *Sweet Dreams. Philosophical Obstacles to a Science of Consciousness*, MIT Press, Cambridge (MA) 2005, cap. 2 (trad. it., *Sweet Dreams. Illusioni filosofiche sulla co-*

scienza, Cortina, Milano 2006); *Heterophenomenology reconsidered*, in «Phenomenology and Cognitive Science», 6(1-2), 2007, pp. 247-270. Per un commento su quest'ultimo saggio di Dennett, vedi S. Vallor, *The fantasy of third-person science: Phenomenology, ontology and evidence*, in «Phenomenology and Cognitive Sciences», *online first*.

11. Una tradizione ben ricostruita in W. Lyons, *The Disappearance of Introspection*, MIT Press, Cambridge (MA) 1986; su cui vedi D.C. Dennett, *Review of William E. Lyons' The Disappearance of Introspection*, in «Philosophy of Science», 55(4), 1988, pp. 653-654.
12. In «Psychological Review», 84, 1977, pp. 231-259.
13. Sugli sviluppi di questa tesi, vedi R.E. Nisbett-L. Ross, *Human Inference: Strategies and Shortcomings of Social Judgment*, Prentice-Hall, Englewood Cliffs 1980 (trad. it., *L'inferenza umana. Strategie e lacune del giudizio sociale*, Il Mulino, Bologna 1989); L. Ross-R.E. Nisbett, *The Person and the Situation: Perspectives of Social Psychology*, McGraw-Hill, New York 1991 (trad. it., *La persona e la situazione*, Il Mulino, Bologna 1998); T.D. Wilson, *Strangers to Ourselves: Discovering the Adaptive Unconscious*, Harvard UP, Cambridge (MA) 2002.
14. Vedi A. Gopnik, *How we know our minds: The illusion of first-person knowledge of intentionality*, in «Behavioral and Brain Sciences», 16, 1993, pp. 1-15, 90-101; D. Wegner, *The Illusion of Conscious Will*, MIT Press, Cambridge (MA) 2002; W. Hirstein, *Brain Fiction: Self-Deception and the Riddle of Confabulation*, MIT Press, Cambridge (MA) 2005.
15. Vedi D. Rosenthal, *Introspection*, in *The MIT Encyclopedia of the Cognitive Sciences*, a cura di R.A. Wilson e F.C. Keil, MIT Press, Cambridge (MA) 1998, p. 420.
16. Vedi E. Schwitzgebel, *No unchallengeable epistemic authority, of any sort, regarding our own conscious experience – Contra Dennett?* in «Phenomenology and Cognitive Sciences», 6, 2007, pp. 107-113; *The unreliability of naive introspection*, in «Philosophical Review», 117(2), 2008, pp. 245-273.
17. Vedi, per esempio, O. Flanagan, *Consciousness Reconsidered*, MIT Press, Cambridge (MA) 1992; J. Searle, *The Rediscovery of the Mind*, MIT Press, Cambridge (MA) 1992 (trad. it. *La riscoperta della mente*, Bollati Boringhieri, Torino 1994).
18. Vedi Searle, *Consciousness, explanatory inversion, and cognitive science*, in «Behavioral and Brain Science», 13, 1990, pp. 585-642; *The Rediscovery of the Mind* cit.; K. Ludwig, *On explaining why things look the way they do*, in *Perception*, a cura di K. Akins, Oxford UP, Oxford 1996. Per una rassegna della letteratura su coscienza e intenzionalità, vedi C. Siewert, *Consciousness and intentionality*, in *The Stanford Encyclopedia of Philosophy*, a cura di E.N. Zalta, <http://plato.stanford.edu/archives/spr2007/entries/consciousness-intentionality/>.
19. Il termine è di N. Block (*Consciousness, Function, and Representation. Collected Papers*, vol. 1, MIT Press, Cambridge [MA] 2007, cap. 9) e designa un costrutto funzionale introdotto dallo psicologo per dar conto della capacità di un sistema di aver accesso ai propri stati interni al fine di espletare compiti cognitivi di alto livello quali l'inferenza e la verbalizzazione.
20. W. Lycan, *Consciousness*, MIT Press, Cambridge (MA) 1987; Id. *Consciousness and Experience*, MIT Press, Cambridge (MA) 1996.
21. Sulle teorie dell'ordine superiore, vedi P. Carruthers, *Higher-order theories of consciousness*, in *Stanford Encyclopedia of Philosophy* cit. <http://plato.stanford.edu/archives/fall2007/entries/consciousness-higher/>.
22. Sulle teorie rappresentazionaliste della coscienza, vedi W. Lycan, *Represent-*

ational theories of consciousness, in *Stanford Encyclopedia of Philosophy* cit., <http://plato.stanford.edu/archives/win2007/entries/consciousness-representational/>.

23. Vedi, per esempio, K. Wilkes, *Is consciousness important?* in «British Journal for the Philosophy of Science», 35, 1984, pp. 223-243; Id., *Yishi, duo, us and consciousness*, in *Consciousness in Contemporary Science*, a cura di A. Marcel e E. Bisiach, Oxford UP, Oxford, 1988.
24. Vedi, per esempio, P.S. Churchland, *Consciousness: The transmutation of a concept*, in «Pacific Philosophical Quarterly», 64, 1983, pp. 80-95.
25. R. van Gulick *Consciousness*, in *Stanford Encyclopedia of Philosophy* cit., <http://plato.stanford.edu/archives/spr2008/entries/consciousness/>, § 9.3. Sull'io, si vedano, oltre al cap. 13 di *Coscienza*, i saggi: *The origins of selves*, in «Cogito», 3, 1989, pp. 163-173; *The self as the center of narrative gravity*, in *Self and Consciousness: Multiple Perspectives*, a cura di F. Kessel, P. Cole e D.L. Johnson, Erlbaum, Hillsdale (NJ) 1992, pp. 103-115; *The self as a responding – and responsible – artifact*, in «Annals New York Academy of Sciences», 1001, 2003, pp. 39-50.
26. D.C. Dennett - M. Kinsbourne, *Time and the observer: The where and when of consciousness in the brain*, in «Behavioral and Brain Sciences», 15(2), 1992, pp. 183-234; *Escape from the Cartesian theater* (response to commentators), in «Behavioral and Brain Sciences», 15(2), 1992, pp. 234-247; *Counting consciousness: None, one, two, or none of the above?* (continuing commentary on *Time and the observer*), in «Behavioral and Brain Sciences», 17(1), 1994, pp. 178-180; *Multiple drafts: An eternal golden braid?* (continuing commentary on *Time and the observer*), in «Behavioral and Brain Sciences», 18(4), 1995, pp. 810-811.
27. Vedi soprattutto *Living on the edge*, in «Inquiry», 36, 1993, pp. 135-159; *Caveat emptor*, in «Consciousness and Cognition», 2(1), 1993, pp. 48-57; *The message is: There is no medium*, in «Philosophy and Phenomenological Research», 53(4), 1993, pp. 889-931; *Back from the drawing board (reply to critics)*, in Dahlbom, *Dennett and his Critics* cit., pp. 203-235; *Get real, reply to my critics*, in «Philosophical Topics», 22, 1994, pp. 505-568; *Real consciousness*, in M. Kamppinen e A. Revonsuo (a cura di), *Consciousness in Philosophy and Cognitive Neuroscience*, LEA, Hillsdale (NJ) 1994, pp. 55-63 (rist. in *Brainchildren. Essays on Designing Minds*, MIT Press, Cambridge [MA], 1998, cap. 7); *With a little help from my friends*, in Ross, Brook e Thompson, *Dennett's Philosophy* cit., pp. 362-382.
28. Vedi *supra*, nota 10.
29. D. Volkov, *The moving target: Multiple drafts or fame in brain?* in *Online Papers on Consciousness*, a cura di D. Chalmers e D. Bourget, <http://consc.net/online/1.4c>.
30. D.C. Dennett, *Self-profile*, in *A Companion to the Philosophy of Mind*, a cura di S. Guttenplan, Blackwell, Oxford 1994, p. 238. Per esempio, la visione cieca, discussa da Dennett alle pp. 359-371 di *Coscienza*, può essere descritta come un caso di dissociazione tra la consapevolezza$_2$ (la prestazione del vedente cieco in un compito visivo è superiore al caso) e la consapevolezza$_1$ (il paziente dichiara di non vedere alcunché).
31. Elton, *Daniel Dennett* cit., cap. 5.
32. Ivi, cap. 6. Sulla distinzione personale/subpersonale, vedi Dennett, *Content and Consciousness* cit., pp. 90-6; Id., *The Intentional Stance* cit., capp. 3 e 4.

33. In «Philosophical Review», 85, 1975, pp. 151-171 (rist. in *Brainstorms* cit., cap. 8).
34. Vedi *supra*, nota 8.
35. D.C. Dennett, *Quining qualia*, in *Consciousness in Contemporary Science*, eds. A. Marcel-E. Bisiach, Oxford UP, Oxford, 1988, pp. 42-77 (trad. it., *Quainare i qualia*, in *Mente e corpo. Dai dilemmi della filosofia alle ipotesi della neuroscienza*, a cura di A. De Palma-G. Pareti, Bollati Boringhieri, Torino 2004).
36. Quasi integralmente basati su D.C. Dennett-M. Kinsbourne, *Time and the observer* cit.
37. *Coscienza*, p. 125.
38. Ivi, p. 146.
39. Vedi i commenti a Dennett-Kinsbourne, *Time and the observer* cit.; K. Korb *Stage effects in the Cartesian theater: Review of 'Consciousness Explained'*, in «Psyche», 1(4), 1993, <http://psyche.cs.monash.edu.au/v1/psyche-1-04-korb.html>; W. Robinson, *Orwell, Stalin, and determinate qualia*, in «Pacific Philosophical Quarterly», 75, 1994, pp. 151-164; W. Seager, *Theories of Consciousness: An Introduction and Assessment*, Routledge, London 1999.
40. Vedi W. Lycan *UnCartesian materialism and Lockean introspection*, in «Behavioral and Brain Sciences», 15, 1992, pp. 216-217; R. Van Gulick, *Time for more alternatives*, in «Behavioral and Brain Sciences» pp. 228-229; Korb, *Stage effects in the Cartesian theater* cit.; Robinson, *Orwell, Stalin, and determinate qualia* cit.; C. McGinn, *Consciousness evaded: Comments on Dennett*, in «Philosophical Perspectives», 9, 1995, pp. 241-249; Seager, *Theories of Consciousness* cit., cap. 5.
41. Vedi N. Block, *Begging the question against phenomenal consciousness*, in «Behavioral and Brain Sciences», 15, 1992, pp. 205-206; Korb, *Stage effects in the Cartesian theater* cit.; Seager, *Theories of Consciousness* cit., cap. 5; G. Braddock, *Eliminativism and indeterminate consciousness*, in «Philosophical Psychology», 15(1), 2002, pp. 37-54; J. Christie-J. Barresi, *Using illusory line motion to differentiate misrepresentation (Stalinesque) and misremembering (Orwellian) accounts of consciousness*, in «Consciousness and Cognition», 11, 2002, pp. 347-365.
42. Block, *Begging the question against phenomenal consciousness* cit., p. 205. Su questo punto, vedi anche van Gulick, *Time for more alternatives* cit.; Seager, *Theories of Consciousness* cit., cap. 4; M. Velmans, *Heterophenomenology versus critical phenomenology*, in «Phenomenology and the Cognitive Sciences», 6(1-2), 2007, pp. 221-230.
43. Dennett-Kinsbourne, *Escape from the Cartesian theater* cit., p. 235.
44. S. Schneider, *Daniel Dennett on the nature of consciousness*, in *The Blackwell Companion to Consciousness*, a cura di M. Velmans e S. Schneider, Blackwell, Oxford 2007, p. 323.
45. L'argomento della conoscenza: *Coscienza*, pp. 443-446; *Sweet Dreams* cit., cap. 5. L'argomento dei *qualia* inversi: *Quining qualia* cit.; *Coscienza*, pp. 433-442; *The message is: There is no medium* cit.; *Instead of qualia*, in Kamppinen-Revonsuo, *Consciousness in Philosophy and Cognitive Neuroscience* cit., pp. 129-140 (rist. in *Brainchildren* cit., cap. 8). L'argomento dei *qualia* assenti: *Coscienza*, pp. 450-452; *The unimagined preposterousness of zombies*, in «Journal of Consciousness Studies», 2(4), 1995, pp. 322-326 (rist. in *Brainchildren* cit., cap. 10); *The zombic hunch: Extinction of an intuition?* in *Philosophy at the New Millenium*, a cura di A. O'Hear, Cambridge UP, Cambridge 2001, pp. 27-43 (rist. in *Sweet Dreams* cit., cap. 1).
46. Qui Schneider rinvia a M. Tye, *Qualia*, in *Stanford Encyclopedia of Philosophy*

cit., <http://plato.stanford.edu/archives/sum2003/entries/qualia/>; e R. van Gulick, *Functionalism and qualia*, in *Blackwell Companion to Consciousness* cit., cap. 30.

47. *Coscienza*, pp. 236-237.
48. Vedi anche D.C. Dennett, *Making tools for thinking*, in *Metarepresentations. A Multidisciplinary Perspective*, Oxford UP, Oxford 2000. Idee simili sono state formulate da D. Bickerton in *Language and Species*, University of Chicago Press, Chicago 1990; e in *Language and Human Behavior*, University of Washington Press, Seattle (WA) 1995.
49. Vedi, per esempio, K. Frankish, *A matter of opinion*, in «Philosophical Psychology», 11(4), 1998, pp. 423-442; Id., *Mind and Supermind*, Cambridge UP, Cambridge 2004, cap. 3; P. Carruthers, *The cognitive functions of language*, in «Behavioral and Brain Sciences», 25(6), 2002, pp. 661-662; F. Ferretti, *Perché non siamo speciali. Mente, linguaggio e natura umana*, Laterza, Roma-Bari 2007, cap. 5.
50. Dennett, *Sweet Dreams* cit., p. 135.
51. Cfr. Dennett-Kinsbourne, *Time and the observer* cit., p. 185. La *scomposizione funzionale ricorsiva* è la strategia per cui ciascuno degli *homunculi* che compongono la mente è a sua volta visto come un complesso di *homunculi* più specializzati e dunque progressivamente più semplici; la scomposizione continua finché non si raggiunge un livello in cui i compiti che gli omunculi devono assolvere sono sufficientemente elementari da poter essere giudicati psicologicamente primitivi (vedi *Coscienza*, pp. 292-293). Si noti che questa idea, originariamente definita «comitato cerebrale», è scartata in *Content and Consciousness* cit., p. 87, ma viene vigorosamente ripresa in *Brainstorms* cit., p. 123.
52. B.J. Baars, *A Cognitive Theory of Consciousness*, Cambridge UP, Cambridge 1988; Id., *In the Theater of Consciousness: The Workspace of the Mind*, Oxford UP, Oxford 1997; Id., *The global workspace theory of consciousness*, in *Blackwell Companion to Consciousness* cit., cap. 18. La teoria è stata recentemente sviluppata in neuroscienza cognitiva: vedi S. Dehaene-L. Naccache, *Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework*, in «Cognition», 79, 2001, pp. 1-37; S. Deheane-J.P. Changeux, *Neural mechanisms for access to consciousness*, in *The Cognitive Neurosciences III*, a cura di M. Gazzaniga, MIT Press, Cambridge (MA) 2004, pp. 1145-1159; M. Shanahan e B.J. Baars, *Applying global workspace theory to the frame problem*, in «Cognition», 98, 2005, pp. 157-176.
53. Baars, *A Cognitive Theory of Consciousness* cit., p. 42.
54. Dennett, *Sweet Dreams* cit., p. 136.
55. Vedi D.C. Dennett, *Bewusstsein hat mehr mit Ruhm als mit Fernsehen zu tun*, in *Die Technik auf dem Weg zur Seele*, a cura di C. Maar-E. Pöppel-T. Christaller, Rowohlt, Reinbek 1996, pp. 61-90.
56. Dennett, *Sweet Dreams* cit., pp. 137-138.
57. S. Blackmore, *Destroying the zombic hunch. Review of 'Sweet Dreams' by Daniel Dennett*, in «Nature», 5, 2005, p. 21; Schneider, *Daniel Dennett on the nature of consciousness* cit., p. 319.
58. Baars, *In the Theater of Consciousness* cit., p. 31.
59. Per un'accurata difesa del connubio tra il modello della «celebrità cerebrale» e la teoria dello spazio di lavoro globale, vedi Volkov, *The moving target* cit.

FONTI DELLE ILLUSTRAZIONI

L'autore ringrazia per l'autorizzazione a riprodurre le illustrazioni:

Figura 2.3: © 1969 Harvey Comics Entertainment, Inc.

Figura 2.4: © 1975 Sidney Harris, «American Scientist».

Figura 4.1: Da *Shakey the Computer*. Nils Nilsson. Copyright © 1984. SRI International. Riproduzione autorizzata.

Figura 4.3: Da *The Thinking Computer: Mind Inside Matter*. Bertram Raphael. Copyright © 1976 by W. H. Freeman and Company. Riproduzione autorizzata.

Figura 5.1: Da *A Brief History of Time*. Copyright © 1988 by Stephen W. Hawking. Bantam, Doubleday, Dell Publishing Group, Inc. Riproduzione autorizzata.

Figure 5.5 e 5.6: Da *Mathematical Games* di Martin Gardner. «Scientific American», December 1981. Riproduzione autorizzata.

Figura 5.7: Da *Foundations of Cyclopean Vision*. Bela Julesz. Copyright © 1971 by University of Chicago Press. Riproduzione autorizzata.

Figura 7.4: Tratta da *MacroComputer*, creato da Steven Barney presso il Tufts Curricolar Software Studio.

Figura 8.1: Da *Speaking: From Intention to Articulation.* Willem J. M. Levelt. Copyright © 1989 by Massachusetts Institute of Technology. Riproduzione autorizzata.

Figura 9.1: Da *The Architecture of Cognition* di John R. Anderson, Cambridge, Mass.: Harvard University Press, Copyright © 1983 by the President and Fellows of Harvard College. Riproduzione autorizzata.

Figura 10.1: Da *Brainstorms* di Daniel Dennett. Copyright © 1978 by Bradford Books, Publishers. Published by MIT Press. Riproduzione autorizzata.

Figura 10.7: Disegno di Gahan Wilson © 1990 by The New Yorker Magazine, Inc. Riproduzione autorizzata.

Figura 11.3: Da *Brain Mechanisms in Sensory Substitution.* Paul Bach-y-Rita. Copyright © 1972 Academic Press, Inc. Riproduzione autorizzata.

Figura 11.4: Da «Nature», vol. 221, pp. 963-964. Copyright © 1969 Macmillan Magazines Ltd. Riproduzione autorizzata.

INDICE ANALITICO

INDICE GENERALE